THE LINEAR COMPLEX OF CONICS*

BY

EARL E. LIBMAN

1. Although the complex of first degree curves, that is, the rectilinear complex, has been thoroughly investigated, very little exists in the literature concerning the space complex of curves of the second degree. It is the purpose of this paper to discuss the linear complex of such curves.

Just what would be a linear complex of curves depends, largely, upon the algebra used to investigate them. In the following, where the algebra of quaternions is used, I have called a complex of curves linear when the essential constants of the curve are each a nonhomogeneous linear function of three independent parameters. This may not, perhaps, be what would be designated as the most general linear system under another algebra.

Any conic in space may be represented as the intersection of a right circular cone and a plane. Such a conic is specified by eight essential scalar constants, three for the plane of the conic, and five for the conic upon it. If these are all functions of three independent parameters, we shall call the system a complex of conics. If the eight constants are linear functions of the three parameters, we shall call the system a linear complex of conics. In the following discussion of the linear complex, the algebra used is quaternions with its usual Hamiltonian symbols.

2. If ω is the vector from the origin to the vertex of the right circular cone, $U\alpha$ the direction of the axis of the cone, $\cos^{-1}e$ the vertical semi-angle, γ the normal vector from the origin upon the plane of the conic, and ϱ a variable vector from the origin to any point on the conic, then the conic is represented by the equations

$$(1)\qquad \begin{aligned} S^2\alpha(\varrho-\omega)-e^2\alpha^2(\varrho-\omega)^2 &= 0 \quad \text{(the generating cone)},\\ S\gamma\varrho-1 &= 0 \quad \text{(the plane of the conic)}. \end{aligned}$$

The eight essential scalar constants of the conic (1) are involved in the three vector constants, ω, γ, and $U\alpha$, three each for the vectors ω and γ and two for the unit vector $U\alpha$. These are to be linear functions of three scalar parameters. Let us take for the parameters the components of ω along any three noncoplanar directions. We may then replace the three

* Presented to the Society, February 28, 1925.

scalar parameters by the single vector parameter ω. Hence, for a linear complex as defined above,

$$\begin{aligned} \alpha &= \theta\omega + \lambda, \\ \gamma &= \varphi\omega + \eta, \end{aligned} \tag{2}$$

where θ and φ are linear vector operators and λ and η are constant vectors.

3. For a given γ, equations (2) show that both ω and α are determined.

A. *Upon each plane in space there is one and only one conic of the complex.*

When the conic passes through a point ρ_0, equations (1) give two conditions upon ω. Since ω is equivalent to three scalar parameters,

B. *There is a single infinity of conics of the complex through every point in space.*

If ρ_0 is a point in space, the values of γ corresponding to the conics through ρ_0 satisfy

$$\begin{aligned} S^2[\theta\varphi^{-1}(\gamma-\eta)+\lambda][\rho_0-\varphi^{-1}(\gamma-\eta)] \\ -e^2[\theta\varphi^{-1}(\gamma-\eta)+\lambda]^2[\rho_0-\varphi^{-1}(\gamma-\eta)]^2 &= 0, \\ S\rho_0\gamma - 1 &= 0, \end{aligned} \tag{3}$$

obtained by substituting for ω in terms of γ in equations (1) and putting $\rho = \rho_0$. This is evidently the tangential equation of the cone enveloped by the planes of the conics that pass through the point ρ_0. It is a cone of class four.

If we eliminate ω from

$$\begin{aligned} S^2(\theta\omega+\lambda)(\rho_0-\omega) - e^2(\theta\omega+\lambda)^2(\rho_0-\omega)^2 &= 0, \\ S(\varphi\omega+\eta)\rho_0 - 1 &= 0, \\ S^2(\theta\omega+\lambda)(\rho-\omega) - e^2(\theta\omega+\lambda)^2(\rho-\omega)^2 &= 0, \\ S(\varphi\omega+\eta)\rho - 1 &= 0, \end{aligned} \tag{4}$$

obtained by substituting (2) in (1), and writing the condition that the conic passes through ρ_0, we have a surface upon which lie all the conics of the complex that pass through the point ρ_0.

C. *Associated with every point in space there is a surface and a cone such that the tangent planes to the cone intersect the surface in the conics of the complex that pass through the point.*

The equation of any line in space passing through the point β and parallel to the vector δ is

$$\rho = \beta + x\delta. \tag{5}$$

If the plane of the conic (1) contains this line, then β is a vector from the origin to a point on this plane, and δ is a vector lying in the plane, and we have

$$S\gamma\beta - 1 = 0, \qquad S\gamma\delta = 0. \tag{6}$$

The line intersects the cone (1) in points corresponding to the roots of

$$x^2 S\delta[\alpha S\alpha - e^2\alpha^2]\delta - 2xS(\beta-\omega)[\alpha S\alpha - e^2\alpha^2]\delta + S(\beta-\omega)[\alpha S\alpha - e^2\alpha^2](\beta-\omega) = 0. \tag{7}$$

If the line is tangent to the cone the roots of (7) must be equal. Hence

$$S^2(\beta-\omega)[\alpha S\alpha - e^2\alpha^2]\delta - S\delta[\alpha S\alpha - e^2\alpha^2]\delta S(\beta-\omega)[\alpha S\alpha - e^2\alpha^2](\beta-\omega) = 0. \tag{8}$$

If the line (5) is tangent to the conic then equations (6) and (8) are satisfied and solving gives six values for ω.

D. *Every straight line in space is touched by six conics of the complex.*

4. In as much as the vector ω is at once the parameter of the complex and the vertex of the generating cone, and therefore every conic of the complex corresponds to a certain position of that vertex, we shall call the point ω the generating point, and shall speak of the corresponding conic as being generated by the generating point.

If the plane of the conic passes through its generating point, the conic will obviously be a line pair, since the plane contains the vertex of the cone, and this is the only case of such degeneracy. In order that the plane of the conic contain its generating point we must have

$$S\gamma\omega - 1 = 0. \tag{9}$$

If in this equation we substitute for γ its value in terms of ω from equations (2), we will obtain an expression in ω which is the locus of the centers of all the degenerate conics of the complex. If, on the other hand, we substitute in this equation for ω its value in terms of γ from equations (2), we obtain an expression in γ which is the envelope of the planes of the degenerate conics of the complex. We will call these two surfaces the first and second singular surfaces respectively. Their expressions, as given below, show them to be quadrics:

$$\begin{aligned} &S\omega\varphi\omega + S\omega\eta - 1 = 0 && \text{(the first singular quadric; point equation of locus of centers of degenerate conics);} \\ &S\gamma\varphi^{-1}\gamma - S\gamma\varphi^{-1}\eta - 1 = 0 && \text{(the second singular quadric; tangential equation of surface enveloped by planes of degenerate conics).} \end{aligned} \tag{10}$$

Note that the two singular quadrics have parallel axes.

E. *The centers of the degenerate conics of the complex lie upon a quadric surface, and their planes envelope another quadric surface. These are the first and second singular quadrics.*

5. The equations (1) will represent any conic through the point ϱ_0 that belongs to the complex if ω satisfies

$$\begin{aligned} S^2(\theta\omega+\lambda)(\varrho_0-\omega)-e^2(\theta\omega+\lambda)^2(\varrho_0-\omega)^2 &= 0, \\ S(\varphi\omega+\eta)\varrho_0-1 &= 0, \end{aligned} \tag{11}$$

obtained by substituting (2) in (1) and putting $\varrho=\varrho_0$. It is at once evident that this is a plane quartic curve which is the locus of the generating points corresponding to all conics through the point ϱ_0.

F. *Associated with every point in space there is a plane quartic curve such that when the generating point describes the curve, the corresponding conics pass through the point. We will call this the generating quartic of the point.*

G. *The conics through a given point are in one to one correspondence with the points of the generating quartic of the given point.*

The generating points of the degenerate conics of the complex that pass through a point lie both upon the generating quartic of the point and upon the first singular quadric.

H. *Through every point in space pass eight degenerate conics of the complex.*

6. Let the plane of the generating quartic of the point ϱ_0 be

$$S\xi\varrho-1=0. \tag{12}$$

Then, since ω is a vector to the quartic, we have also $S\xi\omega-1=0$. Writing the equation of the plane (11) in the form

$$S\omega\frac{\varphi'\varrho_0}{1-S\eta\varrho_0}-1=0, \tag{13}$$

where φ' is the conjugate of φ, comparison shows that

$$\xi=\frac{\varphi'\varrho_0}{1-S\eta\varrho_0}, \tag{14}$$

from which

$$\varrho_0=\frac{\varphi'^{-1}\xi}{1+S\xi\varphi^{-1}\eta}. \tag{15}$$

Evidently there is a one to one linear correspondence between point (coördinate ϱ_0) and plane (coördinate ξ) and we will call the plane of the generating quartic of a point its polar plane; and the point the pole of the plane of its generating quartic with respect to the complex:

$$(16)\qquad \begin{aligned} & S\varrho\frac{\varphi'\varrho_0}{1-S\eta\varrho_0}-1=0 && \text{(polar plane of point } \varrho_0 \\ & && \text{with respect to the complex);} \\ & \frac{\varphi'^{-1}\xi}{1+S\eta\varphi'^{-1}\xi} && \text{(pole of the plane } S\xi\varrho-1=0 \\ & && \text{with respect to the complex).} \end{aligned}$$

The names "pole" and "polar" as used above for point and plane are justified by the similarity between the behavior of this point and plane and that of the pole and polar of a quadric. By means of (16) it can be readily shown that

I. *When a point describes a line its polar plane with respect to the complex rotates about another line*;

J. *When a plane rotates about a line its pole with respect to the complex describes another line.*

7. Let us now consider what will be the loci of the poles that lie upon their polar planes and also of the planes that contain their poles. If the point ϱ_0 lies upon its polar plane (16) we have, putting $\varrho=\varrho_0$ in the first equation of (16),

$$(17)\qquad S\varrho_0\varphi\varrho_0+S\eta\varrho_0-1=0.$$

Comparison of this with equations (10) shows that it is the first singular quadric.

K. *The locus of the points that lie upon their polar planes is the first singular quadric.*

Again if the plane $S\xi\varrho-1=0$ contains its pole (16), then substituting $\varrho=(\varphi'^{-1}\xi)/(1+S\eta\varphi'^{-1}\xi)$ in this equation we get

$$(18)\qquad S\xi\varphi^{-1}\xi-S\xi\varphi^{-1}\eta-1=0,$$

which is evidently the second singular quadric.

L. *The envelope of the planes that contain their poles is the second singular quadric.*

8. It is to be noted that, although the machinery used in this work is metric, yet the results are essentially projective. Moreover, it is evident that the Theorems A, B, C, D and H are independent of the method used to designate the conic (that is, the circular cone and the plane) while Theorems I, J, K, L are called into prominence by this construction.

The writer wishes to express his appreciation of the advice and encouragement he has received in this work from Professors J. B. Shaw and A. B. Coble.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

ON THE WEDDLE SURFACE AND ANALOGOUS LOCI*

BY

ARNOLD EMCH

I. Introduction

Geometrically, the well known quartic, the Weddle surface, may be defined as the locus of the vertices of the quadric cones in the linear three-dimensional system of quadrics through six independent points in space. In a similar manner, the locus of the vertices of cones through seven points, or through the eight base points of the net of quadrics determined by the seven points, is a sextic of genus three, which however is not the most general sextic with this genus.

It is the purpose of this paper to establish analogous loci for cubic cones, and, subsequently, for general n-ic cones. The base points through which the cones pass shall be denoted by $\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_i$, the vertices of the coördinate tetrahedron by A_1, A_2, A_3, A_4, the join of any two base points $\mathfrak{A}_i$ and $\mathfrak{A}_k$ by g_{ik} or simply by g; a surface locus of vertices of cones by W, a curve locus by S.

In case of n-ic cones there is $i = N+1$ ($N = n(n+3)/2$) for a W, $i = N+2$, for an S.

Before taking up the cubic cones it is perhaps well to state briefly some of the properties of the locus of vertices of quadric cones through six points, and through the eight base points of a net of quadrics.

II. The Weddle surface

When $n = 2$, $N+1 = 6$, the order of W is $m = 4$, and we have the Weddle surface as the locus of quadric cones through six points. The number of lines g is 15 which in sets of 5 pass through the six points, the nodes of W. The six points can be arranged in 10 double-triples, each of which determines two planes through the six points, or a degenerate cone whose axis lies on W. Thus, in addition to the 15 g's, there are 10 other lines on W, so that the total number of lines on W is 25, as is well known.

The order of the curve S of vertices of quadric cones through seven points is $s = 6$. But in the case of $n = 2$ we have the circumstance that all quadric cones through seven points pass through an eighth point, so that the S_6 appears as the locus of vertices of quadric cones in a net of quadrics.

* Presented to the Society, December 29, 1923.

The S_6 cuts each of the 28 lines joining every pair of the base points in 2 points. Every sextuple of the base points determines a Weddle surface, so that there are 28 Weddle surfaces through the S_6. The S_6 is also the base curve of the involutory cubic Cremona transformation connected with the eight base points, and the surface of the eighth order formed by the trisecants of the S_6 is the base surface of the transformation. The cubic surfaces corresponding to any two planes pass through S_6, and intersect moreover in a cubic rest curve which corresponds to the line of intersection of the two planes. The S_6 is therefore of genus 3. But this S_6 is not the most general sextic of this genus, since the system of general sextics of genus 3 has the dimension 24 that of our system being $3 \cdot 7 = 21$. For 7 points determine such a sextic and there are $\infty^{3 \cdot 7}$ sextuples. What characterizes our sextic is the fact that it is uniquely determined by the 8 base points of a net of quadrics. The number 21 as the dimension of the system of sextics is also verified as follows. There are ∞^{18} sextuples of points and consequently the same manifold of Weddle surfaces. With a definite Weddle surface fixed, every seventh point determines an S_6 uniquely. Hence on every Weddle surface there are ∞^3 sextics, so that their manifold is ∞^{21}.

In the involutory cubic transformation determined by the octuple of base points, bisecants of the S_6 are transformed into bisecants of S_6, a quartic surface through S_6 into a quartic which passes singly through S_6, because every trisecant of S_6 cuts the original quartic in one point only (outside of S_6). The 28 Weddle surfaces through S_6 are therefore invariant in the transformation.

Now with every Weddle surface is also uniquely associated a Geiser transformation of order 7. As is well known this involutory Cremona transformation may be defined by the net of quadrics through the six nodes of the surface and a seventh point P. Then all quadrics of this net will pass through an eighth point P'. Thus to every point P corresponds a point P' and conversely to P' corresponds in the same transformation P. The Weddle surface is pointwise invariant in this transformation and passes through a twisted cubic on the six nodes. The join of P and P' is a bisecant s of this cubic, so that corresponding pairs on s form an involution with the intersection of s and the Weddle surface (outside of the cubic) as double points. Summing up, we have

THEOREM 1. *There are* 28 *Weddle surfaces through a sixtic* S_6 *determined by the octuple of base points of a net of quadrics. Each of the W's is invariant in the involutory cubic transformation determined by the octuple. There are also* 28 *involutory Geiser transformations each leaving one of the W's, and consequently the* S_6*, pointwise invariant.*

The 28 *Geiser transformations determine an infinite discontinuous group G which leaves the sextic pointwise invariant.*

Denoting by T_{ik} the Geiser transformation associated with the Weddle surface W_{ik} which has the six points obtained from the octuple $A_h (h = 1, 2, \cdots, 8)$ by leaving out A_i and $A_k (h = i, h = k)$, T_{ik} has the property that it permutes A_i and A_k and leaves the remaining 6 points of the octuple unchanged. Thus the 28 T_{ik}'s leave the configuration of the octuple unchanged. Any permutation of the 8 points may be obtained as a product of Geiser transformations T_{ik}. For example, $A_1\, A_2\, A_3\, A_4\, A_5\, A_6\, A_7\, A_8$ is transformed into $A_5\, A_3\, A_1\, A_8\, A_4\, A_2\, A_7\, A_6$ by the transformation $T_{68}\, T_{56}\, T_{45}\, T_{36}\, T_{26}\, T_{13}$. This is obvious since every substitution may be obtained as a product of transpositions. The transformations of G therefore transform the octuple in $8! = 40{,}320$ different ways.

To prove that the transformations of G are infinite in number, consider for example the repeated product $T_{78}\, T_{12}\, T_{78}\, T_{12} \cdots$ and the effect which it has on the Weddle surface W_{12}. If the repeated product at some stage could lead to an identity, W_{12}, which in turn is transformed by $T_{78}, T_{12}, T_{78}, T_{12}, \cdots$ into surfaces of definite orders, would accordingly be transformed into itself. Now, as a matter of fact, T_{78} transforms W_{12} into a surface W'_{12} of order 12 with double points at A_7 and A_8, six-fold points at A_3, A_4, A_5, A_6 and eight-fold points at A_1 and A_2, as can easily be ascertained. T_{12} transforms W'_{12} into a surface W''_{12} of order 28 with eight-fold points at A_1 and A_2, 14-fold points at A_3, A_4, A_5, A_6, 18-fold points at A_7, A_8. Next, T_{78} transforms W''_{12} into W'''_{12}, of order 52; again T_{12} transforms W'''_{12} into $W^{(4)}_{12}$ of order 84, T_{78} transforms $W^{(4)}_{12}$ into $W^{(5)}_{12}$ of order 124, etc. Thus the succession of $1, 2, 3, 4, \cdots, m, \cdots$ transformations (including 1 as the identical one) leads consecutively to surfaces of order $4, 12, 28, 52, 84, 124, \cdots, 4(m^2 - m + 1), \cdots$. From this follows that the repeated products of T_{78} and T_{12} all represent different transformations; and that their number, and hence also the number of transformations of the G, is infinite.

III. W-Surface of vertices of cubic cones

Let $\mathfrak{A}_i (a_i\, b_i\, c_i\, d_i)$, $i = 1, 2, \cdots, 10$, be the ten base points and $P(x_1, x_2, x_3, x_4)$ any point in space distinct from any of the $\mathfrak{A}_i$'s. Assume the coördinate tetrahedron in a general position, and project the points $\mathfrak{A}_i$ from P upon $x_4 = 0$. The coördinates of the projections $\mathfrak{A}'_i$ are

$$\varrho X_i = a_i x_4 - d_i x_1, \quad \varrho Y_i = b_i x_4 - d_i x_2, \quad \varrho Z_i = c_i x_4 - d_i x_3,$$

or, writing cartesian coördinates in space

$$x = \frac{x_1}{x_4}, \qquad y = \frac{x_2}{x_4}, \qquad z = \frac{x_3}{x_4},$$

and putting $d_i = 1$,

$$(1) \qquad \sigma X_i = x - a_i, \qquad \sigma Y_i = y - b_i, \qquad \sigma Z_i = z - c_i.$$

The ten projections $\mathfrak{A}_i'$ lie on a cubic in $x_4 = 0$, when the determinant

$$(2) \quad \begin{vmatrix} X_1^3, & Y_1^3, & Z_1^3, & X_1^2Y_1, & X_1^2Z_1, & Y_1^2X_1, & Y_1^2Z_1, & Z_1^2X_1, & Z_1^2Y_1, & X_1 & Y_1 & Z_1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \\ X_{10}^3, & Y_{10}^3, & Z_{10}^3, & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & X_{10} & Y_{10} & Z_{10} \end{vmatrix} = 0.$$

The expansion of this gives an equation of degree 30 in x, y, z for the locus of vertices (x, y, z) of cubic cones through the ten given points. But this equation is reducible, as is seen by expanding the X, Y, and Z terms in (2), and by subtracting the first row in succession from the 9 remaining rows. In this manner all cubic terms in the 9 rows vanish, so that the degree of the expanded determinant now becomes 21. By repeated division by proper numerical factors and subsequent subtraction the degree may be reduced still further. But the degree of the final irreducible equation may be established by a much less laborious method. The order of the W-surface is equal to the number of its intersections with a generic line. Such a line is $y = 0$, $z = 0$. Substituting this in (2) the first column contains only cubic terms in x, the second and third columns only constants, the fourth and fifth only quadratic terms in x, the sixth only linear terms in x, the seventh constant terms, the eighth linear terms in x, the ninth constant terms, and the tenth linear terms in x. The determinant consequently expands into a polynomial of degree $1 \cdot 3 + 2(3-1) + 3(3-2) = 10$ in x. Consequently the line $y = 0$, $z = 0$ cuts the W-surface in 10 points. When the determinant (2) is expanded by multiplying the elements of the kth row by their corresponding minors, a homogeneous polynomial of degree 3 in $x - a_k$, $y - b_k$, $z - c_k$ is obtained. From this follows that the point (a_k, b_k, c_k) is a triple point of the surface, and that the cubic cone with this point as a vertex, and through the remaining nine points, is the tangential cone to the surface at this point.

Consider any point V of a line g joining any two of the ten points $\mathfrak{A}_i$. This line g and the 9 lines joining V to the 8 remaining points determine a cubic cone uniquely. This is a cubic cone through the 10 points, consequently V, and the line g, lie on the W-surface. These results may be summed up in

THEOREM 2. *The locus of vertices of cubic cones through* 10 *fixed points in space is a surface of order* 10 *with these points as triple-points and with the* 45 *joins of the* 10 *points as single lines on the surface.*

IV. S-Curve of vertices of cubic cones

In addition to the 10 fixed points $\mathfrak{A}_1, \cdots, \mathfrak{A}_{10}$ of the W-surface of order 10 consider another fixed point $\mathfrak{A}_{11}$. The locus of the vertices of cubic cones through these 11 points is a curve S which may be obtained from the intersection of any two W-surfaces associated with two sets of ten points each chosen from the 11 points. Thus, S may be obtained from the intersection of the two surfaces W_{10} and W_{11} associated with $\mathfrak{A}_1 \cdots \mathfrak{A}_9 \mathfrak{A}_{10}$ and $\mathfrak{A}_1 \cdots \mathfrak{A}_9 \mathfrak{A}_{11}$, respectively. Now consider the line l joining $\mathfrak{A}_{10}$ and $\mathfrak{A}_{11}$; it will cut W_{10} in 7 points $V_1, \cdots, V_7$ outside of $\mathfrak{A}_{10}$ which is a triple point of W_{10}. Hence $V_1, \cdots, V_7$ are vertices of cubic cones which pass through all 11 points and are therefore points of S. *Thus, each of the* 55 *joins* l *of the* 11 *points cuts the curve* S *in* 7 *points.* S *itself does not pass through the* 11 *points* $\mathfrak{A}_1, \cdots, \mathfrak{A}_{11}$. This fact may be immediately verified analytically by choosing $\mathfrak{A}_{10}$ at ($a_{10} = 0$, $b_{10} = 0$, $c_{10} = 0$), and $y = 0$, $z = 0$ as the line l. Then (2) reduces to an equation of degree 7 in x.

As W_{10} and W_{11} have the 36 joins g of $\mathfrak{A}_1, \cdots, \mathfrak{A}_9$ in common, their curve of intersection outside of these lines is of order 64. But it is reducible. To show this take the plane p determined by any triple of points out of the nine common to W_{10} and W_{11}, say by $\mathfrak{A}_1$, $\mathfrak{A}_2$, $\mathfrak{A}_3$. Let g_1, g_2, g_3 be the joins of $\mathfrak{A}_2\mathfrak{A}_3$, $\mathfrak{A}_3\mathfrak{A}_1$, $\mathfrak{A}_1\mathfrak{A}_2$, and B_1, B_2, $\cdots$, B_{15} the intersections with p of the 15 joins g of $\mathfrak{A}_4, \mathfrak{A}_5, \cdots, \mathfrak{A}_9$. The plane p cuts W_{10} and W_{11} each in g_1, g_2, g_3 and a curve of order 7 which passes through the 15 B's and singly through $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3$, since these are triple points. Hence the two curves of order 7, outside of the B's and $\mathfrak{A}$'s intersect in $49 - 15 - 3 = 31$ residual points, which lie on the curve S. But S cuts each g_1, g_2, g_3 in 7 points. Consequently p intersects S in $31 + 21 = 52$ points, which is the order of S.

As the curve of intersections of W_{10} and W_{11} outside of the 36 g's is of order 64, of which S is a part, there is a residual curve R of order 12. Since each of the 9 $\mathfrak{A}$'s common to W_{10} and W_{11} is a triple point for both surfaces at which 8 g's, common to W_{10} and W_{11}, are concurrent, the R passes through each of the 9 points. Thus, in the plane p, R passes singly through $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3$, and must cut g_1, g_2, g_3 each in 3 additional points. The curve R does not belong to the locus of vertices of cubic cones through 11 points. To ascertain the geometric meaning of R, consider a pencil $\Phi_1 + \lambda\Phi_2 = 0$ of cubic cones through 8 of the 9 points, say $\mathfrak{A}_1, \cdots, \mathfrak{A}_8$, with $V(x, y, z)$ as a vertex. All cones of the pencil pass through the 8 lines $V\mathfrak{A}_1, \cdots, V\mathfrak{A}_8$, and a ninth fixed line g_9. Intersect this pencil with $x_4 = 0$, as before, and denote the projections by the same letters with accents. Then in order that $\mathfrak{A}_9'$ may coincide with the intersections of g_9 with $x_4 = 0$, on substituting the coördinates of $\mathfrak{A}_9$ in

$\Phi_1' + \lambda \Phi_2' = 0$, there must be simultaneously $\Phi_1' = 0$, $\Phi_2' = 0$. The irreducible part of the intersection of the two surfaces is precisely the curve R. From every point of R the nine points are projected into 9 points of a pencil of cubics. Hence there is a cubic of the pencil through $\mathfrak{A}_{10}'$ and another cubic through $\mathfrak{A}_{11}'$. Consequently every point of R is a vertex of two cubic cones through $\mathfrak{A}_1, \cdots, \mathfrak{A}_9, \mathfrak{A}_{10}$ and $\mathfrak{A}_1, \cdots, \mathfrak{A}_9, \mathfrak{A}_{11}$. The result may be stated in

THEOREM 3. *The locus of vertices of cubic cones through* 11 *points is a curve* S *of order* 52, *which cuts each of the* 55 *joins in* 7 *points.*

There are 55 *W-surfaces through* S. *Any two surfaces of the set intersect in a residual curve* R *of order* 12, *uniquely associated with the* 9 *points* $\mathfrak{A}$ *common to the two W's. The curve* R *is the locus of points from which the nine points are projected into the base points of a pencil of cubics.*

The curve S has no effective singularities, so that its genus may be determined by evaluating the number of its apparent double points from any generic point O, and by subtracting this number from the maximum number of double points of its projection. The sextic cone C through S with O as a vertex cuts W_{11}, outside of S, in a residual curve S^* of order $10 \cdot 52 - 52 = 468$. Now S' cuts W_{10} in 4680 points. Among these points are the $9 \cdot 52 = 468$ points in which the first polar of O with respect to W_{11} cuts S. At such a point the tangents to S and S' at this point lie in the tangent plane to W_{11} at this point. Hence the projection of a point does not result in a double point. The cone C cuts each join g of the nine points $\mathfrak{A}$, associated with S, in 52 points, among which are the 7 intersections of g with S, which do not lie on S^*. The remaining 45 points on each g lie on S^*, and on W_{10}, but not on S. Thus there are $36 \cdot 45 = 1620$ such points, which also are not projected into double points of the projections of S' of S from O. The remaining $4680 - 468 - 1620 = 2592$ intersections of S^* with W_{10} arrange themselves into 1296 couples P, P' which lie similtaneously on S and S^* and on lines through O. In fact, such a point of intersection P^* of S^* with W_{10} lies on W_{11} and W_{10}, and consequently on S. But P^* is the projection from O upon W_{11} of a point P and S, so that P^* is projected back into P, and P also lies on S^*.

The genus of S *is therefore* $\frac{1}{2} \cdot 51 \cdot 50 - 1296 = 1254$.

As a verification, the same method applied to the curve S in case of the Weddle surface yields the genus 3, as is well known.

V. *W*-SURFACES OF VERTICES OF *n*-IC CONES

Following the same method and notation as in the case of cubic cones it is easy to generalize the results for n-ic cones, and it is not necessary to restate the arguments in detail.

The determinant (2) is now of order $N+1$ and expands into a polynomial of degree $n(N+1)$ in x, y, z. If we write the equation of the n-ic in $x_4 = 0$ in the form

$$X^n + X^{n-1}\Phi_1 + X^{n-2}\Phi_2 + \cdots + \Phi_n = 0,$$

in which Φ_i is a homogeneus polynomial of degree i in Y and Z, so that Φ_i contains $i+1$ terms, the determinant

$$(3)\quad \begin{vmatrix} (x-a_1)^n, & (x-a_1)^{n-1}(y-b_1), & (x-a_1)^{n-1}(z-c_1) & , \cdots, (z-c_1)^n \\ \cdot & \cdot & \cdot & , \cdots, \quad \cdot \\ \cdot & \cdot & \cdot & , \cdots, \quad \cdot \\ \cdot & \cdot & \cdot & , \cdots, \quad \cdot \\ (x-a_{N+1})^n, & (x-a_{N+1})^{n-1}(y-b_{N+1}), & (x-a_{N+1})^{n-1}(z-c_{N+1}), & \cdots, (z-c_{N+1})^n \end{vmatrix} = 0$$

contains x^n as the highest power in the first column only, x^{n-1} in the next two, x^{n-2} in the next three, x^{n-k} in $k+1$ consecutive columns. The reduced form of (3) gives the equation of the W-surface. It is evident from (3) that the generic line $y = 0$, $z = 0$ gives for its intersection with W an equation of degree

$$(4)\quad m = 1\cdot n + 2(n-1) + 3(n-2) + \cdots + (n-1)2 + n\cdot 1 = \frac{n(n+1)(n+2)}{6}$$

in x. Hence m is the order of the W-surface. Moreover if we write (3) in the form (2) and expand from the elements of the kth row, it is apparent that $\mathfrak{A}_k$ is an n-fold point of W. Again, every point of a join g of two of the $N+1$ points is a vertex of an n-ic cone through the $N+1$ points. Hence

THEOREM 4. *The locus of vertices of n-ic cones through $\frac{1}{2}n(n+3)+1$ points $\mathfrak{A}_i$ is a W-surface of order $\frac{1}{6}n(n+1)(n+2)$ with the $\mathfrak{A}_i$'s as n-fold points and the joins g as single lines. The tangential cones at the $\mathfrak{A}_i$'s are determined by the N joins g through each $\mathfrak{A}_i$.*

VI. *S*-CURVES

The locus of vertices of n-ic cones through $N+2$ points is an S-curve, through which all the $\frac{1}{2}(N+2)(N+1)$ W-surfaces associated with the $(N+1)$-tuples of the $N+2$ points pass. Let W_{N+1} and W_{N+2} be two of these surfaces with the group of common points $(\mathfrak{A}) = (\mathfrak{A}_1, \mathfrak{A}_2, \cdots, \mathfrak{A}_N)$ and with $\mathfrak{A}_{N+1}$ and $\mathfrak{A}_{N+2}$ individualizing the two surfaces respectively. Their full intersection, to which S belongs, is of order $m^2 = \frac{1}{36}n^2(n+1)^2(n+2)^2$,

from which split off the $\frac{1}{2}N(N-1)$ joins g of $(\mathfrak{A})$. The intersection proper of W_{N+1} and W_{N+2} is therefore a curve T of order

$$t = m^2 - \frac{1}{2}N(N-1) = \frac{1}{72}(2n^6+12n^5+17n^4-30n^3-55n^2+54n).$$

This curve is composed of S and a residual curve R. To determine the order of these curves, the number of points in which S cuts a join g must first be established. Let g be the join of $\mathfrak{A}_{N+1}$ and $\mathfrak{A}_{N+2}$. It cuts W_{N+1} in $m-n$ points V, outside of $\mathfrak{A}_{N+1}$. The n-ic cones with the V's as vertices and through $(\mathfrak{A})$ and $\mathfrak{A}_{N+1}$ also pass through $\mathfrak{A}_{N+2}$; consequently the same V's are also vertices of n-ic cones through $(\mathfrak{A})$ and $\mathfrak{A}_{N+2}$, and lie on W_{N+2}, and hence on S. Thus, S cuts each g in $m-n$ points. Next cut both W_{N+1} and W_{N+2} by the plane p through $\mathfrak{A}_1$, $\mathfrak{A}_2$, $\mathfrak{A}_3$. Neglecting the joins g_1, g_2, g_3, which are common to both surfaces, the residual insersections are curves C_1 and C_2 of order $m-3$, which both pass through the $\frac{1}{2}N(N-1)-3$ intersections B of the joins g of the group $(\mathfrak{A})$ (minus the three lines g_1, g_2, g_3). As W_{N+1} and W_{N+2} have n-tuple points at $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3$, and at each of these have $N-1$ g's in common, the curve T has multiplicities of order $n^2-N+1 = \frac{1}{2}(n-1)(n-2)$ at each $\mathfrak{A}_1, \mathfrak{A}_2, \mathfrak{A}_3$. The multiplicity of each C_1 and C_2 at these points is $n-2$. Hence these points and the points B absorb $\frac{1}{2}N(N-1)-3+3(n-2)^2$ of the $(m-3)^2$ intersections of C_1 and C_2. The residual set of intersections, that of S with p, outside of g_1, g_2, g_3, consists therefore of

$$\mu = (m-3)^2 - \frac{1}{2}(N-3)(N-4) - 3(n-2)^2$$
$$= \frac{1}{72}(2n^6+12n^5+17n^4-102n^3-379n^2+1098n-648).$$

To this number must be added the $3(m-n) = \frac{1}{2}(n^3+3n^2-4n)$ intersections of S with g_1, g_2, g_3. Hence the order of S is

$$s = \frac{1}{72}(2n^6+12n^5+17n^4-66n^3-271n^2+954n-648).$$

The order of R, the residual curve of intersection of W_{N+1} and W_{N+2}, is $r = t-s$, or

$$r = \frac{1}{2}(n^3+6n^2-25n+18) = \frac{1}{2}(n-1)(n-2)(n+9).$$

R is the locus of points from which the N joins of the group $(\mathfrak{A})$ are projected into the base points of a pencil of n-ics.

Thus may be stated

THEOREM 5. *The locus of vertices of n-ic cones through* $\frac{1}{2}n(n+3)+2$ *points* $\mathfrak{A}$ *in space is an S-curve of order*

$$s = \frac{1}{72}(2n^6+12n^5+17n^4-66n^3-271n^2+954n-648),$$

which cuts each of the $\frac{1}{2}(N+1)(N+2)$ *joins of each two of the* $\mathfrak{A}$*'s in* $\frac{1}{2}n(n+1)(n+2)-n$ *points, and does not pass through the* $\mathfrak{A}$*'s. Any two W-surfaces each associated with* $N+1$ *points of the set* $\mathfrak{A}$ *pass through S and intersect moreover in a residual curve R of order*

$$r = \frac{1}{2}(n-1)(n-2)(n+9)$$

which has multiple points of order $\frac{1}{2}(n-1)(n-2)$ *at the points of the group* $(\mathfrak{A})$ *associated with R. R is the locus of points from which the group* $(\mathfrak{A})$ *is projected into the base of a pencil of n-ics.*

The genus of S (and also of R) could easily be determined as in the case of cubic cones and, for the sake of brevity, will be omitted.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

ALGEBRAIC SURFACES WITH REDUCIBLE BITANGENT AND OSCULATING HYPERPLANAR SECTIONS*

BY

MARIA CASTELLANI

The purpose of this investigation is to follow along certain lines the researches of the Italian geometric school on surfaces in hyperspace which have particular systems of reducible hyperplanar sections. In this paper two cases are discussed, first, surfaces which are cut by any bitangent hyperplane along reducible curves, second, surfaces which are cut by any osculating hyperplane along reducible curves. The present investigation may be considered as an extension of work done by Castelnuovo and Enriques. The results are obtained with the aid of algebraic geometry and projective differential geometry.

PART I

1. **Reducible bitangent sections.** In order to study what kind of hyperspatial, algebraic irreducible surfaces F, which are not cones, are cut by a bitangent hyperplane in reducible curves we choose the inductive method. We project such a surface from an arbitrary tangent plane and we obtain a surface cut by any tangent hyperplane in a reducible curve. This surface will be either a surface of Veronese of S_5 or a ruled surface.†

Let us suppose that F is not cut by a tangent plane in a reducible curve, consequently F is not a ruled surface or a surface of Veronese: the space S_r to which F belongs is of dimensions $r > 5$† and the space of two arbitrary tangent planes should be an S_5. We introduce also the hypothesis that such an S_5 has no points of F outside of the two established points.‡

Let us take an arbitrary tangent plane π. From π we project F upon an S_{r-3}, obtaining a surface F' which may be a simple or a multiple projection of F. We can show at once that the second case has to be excluded. In fact if the projection upon F is doubled (the case may be easily

* Presented to the Society, May 2, 1925.

† F. Enriques, *Sulla massima dimensione dei sistemi lineari di curve di dato genere appartenenti ad una superficie algebrica*, Atti della Reale Accademia delle Scienze di Torino, vol. 29 (1893–94), pp. 279–280.

‡ A. Terracini, *Su due problemi concernenti la determinazione di alcune classi di superficie considerate da G. Scorza e da F. Palatini*, Atti della Società dei Naturalisti e Matematici di Modena, ser. 5, vol. 6 (1921–1922).

generalised) the space S_3, which projects F from π, has to meet F in two distinct points Q_1 and Q_2 whose projections upon F' have to be the same. Consequently the straight line through Q_1, Q_2 belongs to the S_3 determined by π and one of the two points. We have ∞^2 pairs of points associated with the ∞^2 points of F. It follows that at least ∞^2 chords of F meet π. But as the projections of Q_1 and Q_2 are the same point Q' of F', the tangent planes π_1 and π_2 respectively in Q_1 and Q_2 at F have to project into the same tangent plane π' at Q' to F', therefore the two S_3 projecting the two points Q_1 and Q_2 coincide. The three tangent planes π, π_1, π_2 belong to the same S_3. When we take as the center of projection π_1 instead of π, the two tangent planes π and π_2 are projected into the same plane π', but a tangent plane cannot have two distinct points of contact, so for the coincidence of the projections of the two planes π and π_2, we have the coincidence of the projections of the two points P and Q_2. Accordingly P and Q_2 give one of the above considered ∞^2 chords of F, but the three points P, Q_1, Q_2 cannot be on the same straight line because each chord cannot cut the surface in three points.* Therefore they give a plane σ. The plane σ belongs to the S_3 which projects Q_1 from π, and to the S_3 which projects P from π_1. Evidently the planes π, π_1, π_2 cut σ in straight lines. Accordingly two tangent planes have a common point.

We arrive at the same conclusions if we take another pair of points, that are projected into the same point π_1, consequently two arbitrary tangent planes have a common point, and F is either a surface of Veronese of S_5, or a ruled surface contrary to our hypothesis.† Therefore we can conclude that F is projected from any tangent plane in a one-to-one way upon a surface F' of S_{r-3}, and accordingly the hyperplanar sections of F' are in one-to-one correspondence with the hyperplanar tangent sections of F.

So when the hyperplanar sections of F have genus p, the hyperplanar sections of F' have genus $p-1$ because they correspond to the hyperplanar tangent sections of F. Moreover the hyperplanar tangent sections of F' correspond to the hyperplanar bitangent sections of F, and the former have to be reducible, because the latter are reducible.

Accordingly F' is either a ruled surface or a surface of Veronese of S_5. We shall study the two cases separately.

2. **The surface F' is a ruled surface.** As the surface F' has ∞^1 straight lines the surface F has a pencil (C) of rational curves C cut in one variable

* E. Bertini, *Introduzione alla Geometria Proiettiva degli Iperspazi*, Pisa, Spoerri, 1907, pp. 194-195.

† P. Del Pezzo, *Sulle superficie dell' n^{mo} ordine immerse nello spazio di n dimensioni*, Rendiconti del Circolo Matematico di Palermo, vol. 1 (1887), pp. 241-271.

point by the S_{r-1} of S_r because the pencil (C) is projected on ∞^1 straight lines of F'.

The pencil (C) may be irrational or rational.

(a) Let us suppose (C) irrational.

The pencil (C) cannot change with the center P of projection, because if (C) changes there will be a continuum of irrational pencils and consequently upon each section of F a continuum of irrational involutions. This is impossible because of a theorem of Castelnuovo.*

If the pencil (C) is not variable any hyperplane through π cuts any C in several points, but under our assumption only one is variable, so apart from this one variable point, there are some fixed points on the plane π. The tangent plane π has to meet all the curves C, and similarly for all the tangents S_2. But as the tangent planes do not cut F in curves it follows that the tangent planes have to pass through at least one basis point of the infinite pencil (C).† Accordingly the projection of F from an S_{r-4} upon an S_3 should be a surface with all tangent planes having a point in common, therefore this surface is a cone and also F is a cone contrary to our hypothesis.

Then F is a surface with an ∞^1 of curves C and the hyperplane through any tangent plane which does not meet them cuts them in only one point. Hence F is a ruled surface.

(b) The pencil is rational.

The surface F to which a rational pencil of rational curves belongs is rational according to a well known theorem of Noether.‡ Such a surface can be represented upon a plane with a system of curves in one-to-one correspondence with hyperplanar sections.

So, as we can show from the plane representation, any S_{r-1} through π (an arbitrary tangent plane) cutting C in only one variable point, cuts F in rational irreducible curves. Therefore $p-1=0$ and $p=1$.

If the surface F is a surface with the hyperplanar sections of genus p then, from a theorem of Castelnuovo§ F has order $r \leq 9$ and can be represented upon a plane by the system of cubics with $9-r$ simple points in common, or with the system of quartics with 2 double common points.

* G. Castelnuovo, *Sulla linearità delle involuzioni piu volte infinite appartenenti ad una curva algebraica,* Atti della Reale Accademia delle Scienze di Torino, vol. 28 (1892–93), pp. 727–738.

† G. Castelnuovo, *Massima dimensione dei sistemi lineari di curve piane di dato genere,* Annali di Matematica, ser. 2, vol. 18 (1890), pp. 119–128.

‡ M. Noether, *Ueber Flächen, welche Schaaren rationaler Curven besitzen,* Mathematische Annalen, vol. 3 (1871), pp. 161–270.

§ G. Castelnuovo, *Sulle superficie algebriche le cui sezioni piane sone curve ellittiche,* Rendiconti della Reale Accademia dei Lincei, ser. 4, vol. 3 (1894), pp. 59–61.

But here we take only the first case into consideration, because the second gives a surface of the eighth order and second kind of S_8, which is projected by a tangent plane not into a ruled surface but into a surface of Veronese.

The cubics corresponding to the hyperplanar bitangent sections are the cubics with two common double points outside the common points of the system. Therefore the cubics are composed of conics of a system ∞^{r-6} and with one common straight line joining the two points corresponding to the two points of contact of the bitangent hyperplanes.

When $r = 9$ the hyperplanar bitangent sections of a surface F are composed of curves of the third order and the sixth order. The first are common to all ∞^3 bitangent S_8 at a fixed point of F, the second are variable in an ∞^3 system with two base points.

3. **Surface F' a surface of Veronese of S_5.** Evidently as $r-3=5$, the surface F belongs to an S_8 and it is a rational surface because it is in one-to-one correspondence with the surface of Veronese, and therefore F can be represented upon a plane.

As in the above mentioned case, we arrive at the conclusion that the tangent sections of F have genus $p-1=0$, because they are in correspondence with the hyperplanar rational sections of the surface of Veronese.

So also in this case the hyperplanar sections of F have genus $p = 1$.

Evidently F is a surface of the eighth order and second kind in correspondence with the system of plane quartics with two common double points.*

The plane curves corresponding to the hyperplanar tangent sections of F are quartics with two other double points. These quartics are composed of two conics of a pencil. Accordingly the bitangent sections of the S_{r-1} cut F in a pencil of curves of the fourth order of which the two points of contact are basis points.

THEOREM I. *The algebraic irreducible surfaces of $S_r (r>5)$, not cones, which any bitangent hyperplane cuts in reducible curves are the ruled surfaces and the surfaces with hyperplanar sections of genus $p = 1$.*

PART II

1. **Reducible osculating sections.**† We shall consider only the fundamental types of surfaces F, because many difficulties arise when we leave the neighborhood of the first order of a point and enter into the neighborhoods of a higher order. So we make the following assumptions:

* Del Pezzo, *Sulle superficie dell' n^{mo} ordine immerse nello spazio di n dimensioni*, Rendiconti del Circolo Matematico di Palermo, vol. 1 (1887), pp. 241-271.

† The osculating sections are given by the hyperplanes through the minimum osculating space. If we take the notation of Bompiani we represent this as $S(2)$. This means that we have to take the derived points of the first and second order of a fixed point. All the planes osculating the curves through a fixed point lie in an $S(2)$. The $S(2)$ is in general

(*a*) F is an algebraic irreducible surface, not a cone, of a convenient S_r; (*b*) the osculating S_{r-1} have to cut F in curves with an ordinary triple point at the point of osculation;* apart from such a point they may not necessarily contain another point of osculation or any point of contact† and may not have a common space; (*c*) there are ∞^2 $S(2)$—one for each point of the surface—the $S(2)$ are S_5 and they do not cut the surface in reducible curves when they cut F; (*d*) two tangent planes at two consecutive points give an S_4; these S_5 do not cut the surface in curves and do not have a common space.

The first type of surface which satisfies these conditions is the surface with reducible tangent sections, hence for $r > 5$ the ruled surfaces.‡

If we consider the osculating S_{r-1} at any non-singular point P of F these osculating S_{r-1} have for definition a common $S(2)$. With our hypothesis this common space is S_5, accordingly such S_r form a linear system ∞^{r-6} and in such a system the curves which they cut on F are variable.§

But for our purpose such curves have to be reducible. So, by a theorem of Enriques, either they are composed of several curves of a rational or irrational pencil of which only one goes through an arbitrary point of F, or they are composed of common curves and others, reducible or not, of a system ∞^{r-6}.

Let us study the two cases separately.

2. **Osculating sections at one point are composed of curves of a pencil.** The osculating hyperplanes at a point P of F cut the surface in curves composed of some curves of a pencil (C). We shall show that the pencil (C) varies with P, that all the pencils thus obtained lie in a system ∞^2 and that the curves C are of the third order and therefore the dimension r of the space is $\leqq 9$.

In fact if (C) does not vary with P, either all the osculating sections have common fixed curves, which is contrary to our hypothesis, or two

an S_5 (Bompiani, *Sopra alcune estensioni dei teoremi die Meusnier e di Eulero*, Atti della Reale Accademia delle Scienze di Torino, vol. 48 (1912–1913), p. 393). The osculating hyperplane cuts a surface in a curve with at least a triple point at the point of osculation.

* Del Pezzo, *Sulle superficie dell' n^{mo} ordine immerse nello spazio di n dimensioni*, Rendiconti del Circolo Matematico di Palermo, vol. 1 (1887), pp. 241–271.

† E. Bompiani, *Determinazione delle superficie integrali di un sistema di equazioni a derivate parziali lineari ed omogenee*, Atti della Reale Accademia delle Scienze di Torino, vol. 52 (1919–1920), pp. 614–615.

M. Castellani, *Sulle superficie i cui spazi osculatori sono biosculatori*, Rendiconti della Reale Accademia dei Lincei, ser. 4, vol. 29 (1922), pp. 347–350.

‡ F. Enriques, *Sulla massima dimensione dei sistemi lineari di curve di dato genere appartenenti ad una superficie algebrica*, Atti della Reale Accademia delle Scienze di Torino, vol. 29 (1893–1894), pp. 279–280.

§ E. Bertini, *Introduzione alla Geometria Proiettiva degli Iperspazi*, p. 26.

curves C go through each point of the surface, which is contrary to the theorem of Enriques, and therefore the pencil P varies with the point P. When the point P varies the pencil (C) gives a continuous system at least ∞^2.

Such a system cannot have basis points, because by our hypothesis the osculating hyperplanes cannot have a common space, therefore the curves C cut each other in ν variable points.

If $\nu \geqq 2$ every pencil (C) would have $\nu \geqq 2$ basis points, coincident or not, and the osculating S_{r-1} at P of F would have points of contact or osculation not at P, or a point of osculation of higher order in P. But such a case is contrary to our hypothesis, so we may conclude that the curves C of (C) cut each other only in P where they have a simple intersection.

The curves C of the system described by the pencil (C) cut each other only in one variable point, accordingly through any two points of F only one curve of the system passes. By a theorem of Enriques* it is a linear system ∞^2. The curves C are rational because the characteristic series on any of them is a g_1^1; therefore it is easy to show that the curves C are rational curves of the third order of S_3.

In fact let us take an arbitrary curve C and an S_{r-1} which osculates the surface in a point not belonging to C. Such an S_{r-1} cuts the surface in three curves and as each of these meets C in one point only, the S_{r-1} intersects C in three points and accordingly the curves C are curves of the third order. Since they are rational and not plane, they are cubics of S_3.†

The surface F to which an ∞^2 of curves of the third order belong is the rational, normal surface of S_9 which can be represented on a plane by the system of plane cubics—or a projection.

3. **Osculating sections at one point have a common curve.** Let us now study the case in which the osculating hyperplanes at a non-singular point P of a surface F cut the surface in a fixed irreducible curve T and in curves C of a system.

The fixed curve T evidently belongs to the $S(2)$ osculating the surface F at the point P. Therefore our problem is now to study the maximum space of curves that the ∞^2 osculating $S(2)$ cut upon a surface.

We have to take into consideration two cases: (a) there are ∞^1 curves on F; (b) there are ∞^2 curves on F.

---

* F. Enriques, *Ricerche di geometria sopra una superficie algebrica*, Memorie della Reale Accademia delle Scienze di Torino, ser. 2, vol. 44 (1894), pp. 179–232.

† A surface with ∞^n curves of S_n belongs to an S_r where $r \leqq n(n+3)/2$; G. Castelnuovo, *Sulle superficie algebriche le cui sezioni piane sono curve ellittiche*, Rendiconti della Reale Accademia dei Lincei, ser. 4, vol. 3 (1894), pp. 59–61; E. Bompiani, *Proprietà differenziali caratteristiche di enti algebrici*, Memorie della Reale Accademia dei Lincei, ser. 4, vol. 29 (1922).

(*a*) Let us begin by supposing that they are curves of S_5. In such a case the osculating $S(2)$ at any point of T are the S_5 of T, so such an S_5 osculates T of the surface F at all its points. Accordingly in this space there will be T and two other curves T' and T'' infinitely near to T'. If we repeat the same process to T' we find that in the S_5 of T there will be T', T'', T''' and in the same way we can conclude that the T' belong to such an S_5 and therefore the entire surface F.

When the curves T' belong to an S_4 it is easy to find a similar result. In fact the S of T' and the tangent at the point P'—infinitely near P—to its curve T' are by definition in the same $S_5(2)$—osculating F in P. Then when P describes T' we find that the S_4 of T' is met by all tangents at any point of T' because P' describes T'. Consequently T and T' are in the same S_5, but such an S_5 may be fixed with the S_4 of T and any point of T'. The same property holds for T' and T''; as a point of T''' is in the osculating $S(2)$ at P we find that this $S(2)$ osculates T' in all its points. As in the case just considered we find that F lies in such an S_5. When the F lies in S_3 we have several cases.

When two S_3 infinitely near together belong to an S_7, as any $S(2)$ osculating F at any point of T may be given by the S_3 of T and a tangent at T' in the point P' infinitely near to P not on T we have the S_7 of two infinitely near curves which osculate F at any point of T, and consequently F is a surface of S_7. We can arrive at the same results if we suppose two curves T and T' in S_6 or S_5.

If instead two curves are in an S_4, by the theorem of Bompiani* it follows that F is a surface with ∞^1 curves T in S_3 of a developable surface. Along such curves the S_5 osculate T in all points. The T in S_3 gives results contrary to our fundamental hypothesis.

The case of plane curves is always possible.

(*b*) When the T are ∞^2 the S_r have as dimensions $r < 8$. In order to show this property we take into consideration some results of algebraic geometry.

Let us take an arbitrary point P of T. The $S(2)$ osculating the surface at P cuts F in a curve T. Any S_{r-1} through such an $S(2)$ cuts the surface in T and in another curve C of a system ∞^{r-6}. By our previous hypothesis this section does not have a point of contact or osculation except at the fixed point P and the curve C will not meet T in a variable point. Accordingly when we take the curves C for any point of T, any one of C is reduced to T and in general to another curve of a system ∞^{r-7}. The existence of such a residual curve is not needed, only we have to observe

* E. Bompiani, *Determinazione delle superficie integrali di un sistema di equazioni a derivate parziali lineari ed omogenee*, Atti della Reale Accademia delle Scienze di Torino, vol. 52 (1919–1920), pp. 614–615.

that from the system of the $\infty^{r-6} S_{r-1}$ osculating at P we have taken a system ∞^{r-7} of hyperplanes tangential to F at every point of T. But by a well known theorem* a linear system ∞^{r-7} of hyperplanes has an S_6' in common, so the S_6' common to all the above mentioned hyperplanes is tangent to F at every point of T. If the curve C by the condition imposed is reduced to two curves like T we have the S_6 osculating F in all points T so in such an S_6 it would be the ∞^1 T and the entire surface would belong to an S_6.

Now let us consider the first case and take such an S_6 and an $S(2)$ osculating F at any point of T, for instance at Q. In such an $S(2)$ there are two tangent planes at two points of the curve T infinitely near together and by our hypothesis two infinitely near tangent planes give an S_4. The S_6 above considered osculates F at every point of T. We observe that the curves T are ∞^2 and we have ∞^1 curves T for any point of F, so the curves T meet each other in at least one variable point.

Any $S(2)$ osculating F in a point other than the points P and Q has in common with this S_7 the S_4 of two tangent planes at two consecutive points of T and the point of intersection R of curves cut by the $S(2)$ osculating Q. Also the $S(2)$ that we have now taken in consideration belongs to such an S_7. R does not belong to the S_4 of two consecutive tangent planes above considered because if this were so it would also be true for any S_4 tangent to two consecutive points of T. This means that either all the S_4 have a common point or as P describes T, there are ∞^1 R. Both cases are contrary to our hypothesis. So the point R and such an S_4 give an $S_5(2)$ osculating F in a point of T. Accordingly such an $S(2)$ is in the S_7 above considered. Consequently the S_7 contains all the $S(2)$ osculating T at every point Q, so that it contains ∞^1 curves and obviously the entire surface.

THEOREM II. *Two classes of algebraic surfaces (not cones) are cut in reducible curves by the osculating S_{r-1} when the S_{r-1} have neither a common space nor points of osculation outside the fixed point; when the $S(2)$ are $\infty^2 S_5$ which do not cut the surface in reducible curves and when the S_4 of two consecutive tangent planes do not cut the surface:*

(1) *the surfaces of $S_r (r > 7)$ cut in curves by the $S(2)$ are ruled surfaces and surfaces with ∞^1 plane curves;*

(2) *the surfaces of $S_r (r > 6)$ which the $S(2)$ do not cut are the rational normal surfaces of S_9 with ∞^2 curves of the third order or a projection.*

---

* E. Bertini, *Introduzione alla Geometria Proiettiva degli Iperspazi*, p. 26.

BRYN MAWR COLLEGE,
BRYN MAWR, PA.

POLYNOMIALS OF SEVERAL VARIABLES AND THEIR RESIDUE SYSTEMS*

BY

AUBREY J. KEMPNER

INTRODUCTION

This paper gives an extension to polynomials of more than one variable of the theory of residue systems of a polynomial of a single variable.† The methods employed are similar to those of the earlier paper, which we shall therefore use freely and refer to as R. I.

All letters involved denote rational integers.

We recall the following definitions and facts.

A polynomial $f(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_n$ was called residually congruent zero modulo m, $f(x) \equiv\equiv 0 \pmod m$, when for all rational integral values of x we have $f(x) \equiv 0 \pmod m$. Similarly $f_1(x) \equiv\equiv f_2(x) \pmod m$, for two polynomials $f_1(x)$, $f_2(x)$, was defined.

For a positive integer m, $\mu(m)$ or μ_m denoted the smallest positive integer such that $\mu_m! \equiv 0 \pmod m$.

For a given modulus m we constructed, by a simple arithmetical process, the *signature* of m, $S(m)$ which depends on m alone,

$$S(m) = \begin{pmatrix} \mu(m) & \mu(d_1) & \cdots & \mu(d_i) & \cdots & \mu(d_{\tau-1}) & \mu(d_\tau) = 0 \\ m/d_0 = 1 & m/d_1 & \cdots & m/d_i & \cdots & m/d_{\tau-1} & m/d_\tau = m \end{pmatrix},$$

where each d_i is a certain proper factor of all d_j, $j > i$, and consequently a divisor of m, with $d_0 = m$, $d_\tau = 1$.

From $S(m)$ we could write down $C(m)$, the *characteristic* of m,

$$C(m) = \begin{pmatrix} \mu(d_{\tau-1}) & \mu(d_{\tau-2}) - \mu(d_{\tau-1}) & \mu(d_{\tau-3}) - \mu(d_{\tau-2}) & \cdots & \mu(m) - \mu(d_1) \\ m & m/d_{\tau-1} & m/d_{\tau-2} & & m/d_1 \end{pmatrix}.$$

$S(m)$ completely determines a chain of residual congruences, modulo m,

$$\frac{m}{d_i} \prod_{j=0}^{\mu(d_i)} (x-j) \equiv\equiv 0 \pmod m \qquad (i = 0, 1, \cdots, \tau),‡$$

* Presented to the Society, December 26, 1924.

† Kempner, these Transactions, vol. 22 (1921), pp. 1–48.

‡ Or $(i = 0, 1, \cdots, \tau - 1)$, the congruence corresponding to $i = \tau$ being trivial.

that is, a set of congruences of the form

$$\frac{m}{d_i} \cdot x^{\mu(d_i)} \equiv \psi(x) \pmod{m},$$

where $\psi(x)$ is a polynomial of degree $< \mu(d_i)$ with integral coefficients in more condensed notation, simply $\{\mu(d_i), m/d_i\}$. We also saw that it was sufficient, in considering residual congruences $a_0 x^k + a_1 x^{k-1} + \cdots + a_k \equiv 0 \pmod{m}$, to restrict a_0 to the divisors of m, including $a_0 = 1$ and $a_0 = m$.

Finally, this chain of congruences has the property that if there is any other residual congruence $cx^\nu \equiv \varphi(x) \pmod{m}$, where c is a divisor of m, then there is in our chain always a congruence for which either $c = m/d_i$, $\nu = \mu(d_i)$, or $c > m/d_i$, $\nu = \mu(d_i)$, or $c = m/d_i$, $\nu > \mu(d_i)$, or $c > m/d_i$, $\nu > \mu(d_i)$. We expressed this fact by saying that for every possible residual congruence modulo m there is in the chain a congruence which is not weaker than it. Its meaning is that every other residual congruence modulo m is a consequence of the congruences of the chain.

Reduced polynomials

1. **Definition.** *We call a polynomial $f(x_1, \cdots, x_k)$ residually congruent modulo m to another polynomial $\varphi(x_1, \cdots, x_k)$, $f \equiv \varphi \pmod{m}$, when for all sets of rational integral values of $x_1, \cdots, x_k$ we have $f \equiv \varphi \pmod{m}$.*

Lemma I. *For a given modulus m and any number k of variables, there exist polynomials $f(x_1, \cdots, x_k) \equiv 0 \pmod{m}$, f of the form $1 \cdot x_1^{\mu(m)} + \cdots + 1 \cdot x_k^{\mu(m)} + \varphi(x_1, \cdots, x_k)$, where φ is of degree $< \mu_m$ in each variable.*

Proof. For one variable, compare R. I, Lemma V. Let $f(x)$ be such a polynomial, then consider, for example, $f(x_1) + \cdots + f(x_k)$.

Our process of reduction of polynomials of more than one variable is carried out in the following manner.

For a given m, and k variables, we shall obviously have the following residual congruences, at least: For x_1, the τ congruences

$$\frac{m}{d_i} \cdot \prod_{j=0}^{\mu(d_i)-1} (x_1 - j) \equiv 0 \pmod{m} \quad (i = 0, 1, \cdots, \tau - 1)$$

or, as we also write, $\{\mu(d_i), m/d_i\}$ or $\{\mu(d_i), m/d_i\}_{x_1}$. Similarly, for each x_ν, $\{\mu(d_i), m/d_i\}_{x_\nu}$.

We first show how far a given polynomial in $x_1, \cdots, x_k$ may be reduced by these congruences; then we shall show that no further reduction is

possible, thus proving that all other conceivable residual congruences modulo m are implied by—that is, are consequences of—the congruences

$$\text{I} \qquad \{\mu(d_i),\ m/d_i\}_{x_\nu} \qquad (i = 0, 1, \cdots, \tau - 1;\ \nu = 1, \cdots, k).$$

For this reason we shall call the $k \cdot \tau$ congruences I a chain of residual congruences modulo m for k variables.

Assume first in our polynomial a term of type $cx_1^{\alpha_1} \cdots x_k^{\alpha_k}$, where at least one of the exponents $\alpha_1, \cdots, \alpha_k \geqq \mu(m)$. Assume, for example, α_1 to be the largest of these exponents, in case there is a largest one, or one of the largest, if two or more are equal. We then depress $cx_1^{\alpha_1} \cdots x_k^{\alpha_k}$ by applying $\{\mu(m),\ m/m = 1\}_{x_1}$, and similarly for each exponent which is $\geqq \mu(m)$.

We thereby obtain as a first step the following reduction.

LEMMA II. *For any given polynomial $f(x_1, \cdots, x_k)$ there exists at least one polynomial $g(x_1, \cdots, x_k)$ satisfying the conditions*

(1) $f \equiv g \pmod m$;

(2) *g is of degree at most $\mu_m - 1$ in each variable separately, and consequently of total degree at most $k \cdot \mu_m - k$;*

(3) *each coefficient of g has one of the values $0, 1, \cdots, m-1$.*

This is accomplished by the sole use of the congruences $\{\mu(m), 1\}$, written for $x_1, \cdots, x_k$ separately. We examine next the influence of the remaining congruences $\{\mu(d_i),\ m/d_i\}_{x_j}$, $i > 0$.

Our polynomial has now no term with exponents $\geqq \mu(m)$. We select a term, if there be any such, in which there is at least one exponent λ, $\mu(m) > \lambda \geqq \mu(d_1)$. Assume for example β_1 in $ex_1^{\beta_1} \cdots x_k^{\beta_k}$ to be such an exponent. Using $\{\mu(d_1),\ m/d_1\}_{x_1}$, we reduce all such powers of x_1 to the power $x_1^{\mu(d_1)}$ and lower powers, and the coefficient e to one of the values $0, 1, \cdots, m/d_1 - 1$. Similarly we reduce all powers of $x_2, \cdots, x_k$ which have exponents between $\mu(m)$ (exclusive) and $\mu(d_1)$ (inclusive). We continue in the same manner for all exponents between $\mu(d_1)$ (exclusive) and $\mu(d_2)$ (inclusive), using the congruence $\{\mu(d_2),\ m/d_2\}_{x_j}$; etc. We shall finally reach the following result (compare R. I, § 5):

THEOREM I. *Assume any polynomial $f(x_1, \cdots, x_k)$. Then there exists at least one polynomial*

$$g(x_1, \cdots, x_k) = \sum_{i_k=0}^{\lambda_k} \cdots \sum_{i_1=0}^{\lambda_1} a_{i_1 i_2 \cdots i_k} x_1^{i_1} \cdots x_k^{i_k}$$

for which

(1) $f \equiv g \pmod m$,

(2) $\lambda_j \leqq \mu(m) - 1 \qquad (j = 1, \cdots, k)$,

(3) $a_{i_1 \cdots i_k}$ $(i_1 = 0, 1, \cdots, \mu(d_{\tau-1}) - 1; \cdots; i_k = 0, 1, \cdots, \mu(d_{\tau-1}) - 1)$

has a value $0, 1, \cdots, m-1$,

(4) $a_{i_1 \cdots i_k}$ $(i_1, \cdots, i_k = \mu(d_{\tau-1}), \cdots, \mu(d_{\tau-2}) - 1)$

has a value $0, 1, \cdots, m/d_{\tau-1} - 1$,

. . . . . .

(5) $a_{i_1 \cdots i_k}$ $(i_1, \cdots, i_k = \mu(d_1), \cdots, \mu(m) - 1)$

has a value $0, 1, \cdots, m/d_1 - 1$.

2. **Definition.** *Polynomials satisfying* (2), (3), (4), (5) *of Theorem* I *we call completely reduced modulo* m.

We wish to show that our reduction is complete in this sense:

THEOREM II. *Every polynomial is residually congruent modulo* m *to exactly one completely reduced polynomial.*

Before entering upon the proof, we recall that we have in the proof of Theorem I exactly exhausted the force of the congruences I. Therefore in agreement with our definition in R. I (§ 3, and § 4, Theorem III) the uniqueness theorem just stated is equivalent to

THEOREM III. *The congruences* I *form a chain of residual congruences modulo* m *for* k *variables.*

Proof. Our theorem is obviously of the same content as the following statement: If $g_1 = \sum \cdots \sum a_{i_1 \cdots i_k} x_1^{i_1} \cdots x_k^{i_k}$ and $g_2 = \sum \cdots \sum b_{i_1 \cdots i_k} x_1^{i_1} \cdots x_k^{i_k}$ are two completely reduced polynomials modulo m, then $g_1 \equiv g_2$ when and only when $a_{i_1 \cdots i_k} = b_{i_1 \cdots i_k}$ for all $i_1, \cdots, i_k$, that is, $g = g_1 - g_2$ is $\equiv 0$ only when all coefficients vanish.

If there should exist any additional residual congruence modulo m which is not implied by the congruences I, then the application of these congruences to it will show that it may be reduced to the form $f(x_1, \cdots, x_k) \equiv 0 \pmod{m}$, where

(1) the highest degree of any x_j which occurs is exactly one of the $\mu(d_i)$,

(2) each term containing this $x_j^{\mu(d_i)}$ will have its coefficient $= 0, 1, \cdots, m/d_i - 1$.

These properties are suffiicient to enable us to argue precisely as in R. I for polynomials of a single variable. Pages 9, 10 of the former paper carry over with only natural modifications, and need not be reprinted. Theorem II and III are thus proved.

3. For the case of two variables, x_1, x_2, a geometrical scheme for exhibiting the character of the completely reduced polynomials is easily devised:

We arrange the coefficients of any completely reduced polynomial modulo m, $f = \sum \sum a_{ij} x_1^i x_2^j$ in a rectangular array

$$
\begin{array}{cccc}
a_{00} & a_{10} & \cdots & a_{\mu(m)-1,\,0} \\
a_{01} & a_{11} & \cdots & a_{\mu(m)-1,\,1} \\
\cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot \\
\cdot & \cdot & \cdots & \cdot \\
a_{0,\,\mu(m)-1} & a_{1,\,\mu(m)-1} & \cdots & a_{\mu(m)-1,\,\mu(m)-1}.
\end{array}
$$

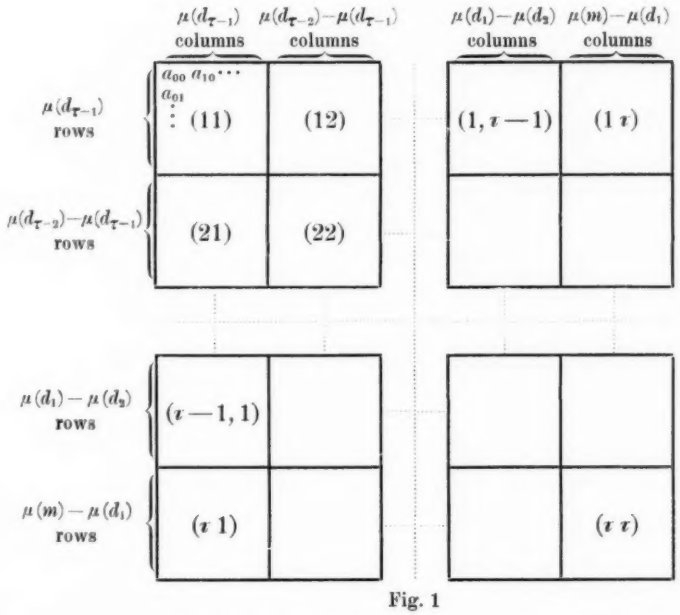


Fig. 1

If we denote by (s, t) the sth horizontal and tth vertical of the rectangular blocks into which our array is broken up, then each of the $[\mu(d_{\tau-1})]^2$ coefficients a_{ij} lying in $(1, 1)$ has one of the m values $0, 1, \cdots, m-1$; each of the $\mu^2 d_{\tau-2} - \mu^2 d_{\tau-1}$* coefficients in $(1, 2)+(2, 2)+(2, 1)$ has one of the $m/d_{\tau-1}$ values $0, 1, \cdots, m/d_{\tau-1}-1$; each of the $\mu^2 d_{\tau-3} - \mu^2 d_{\tau-2}$ coefficients in $(1, 3)+(2, 3)+(3, 3)+(3, 2)+(3, 1)$ has one of the $m/d_{\tau-2}$ values $0, 1, \cdots, m/d_{\tau-2}-1$, etc.; finally, each of the $\mu^2 m - \mu^2 d_1$ coefficients in $(1, \tau)+(2, \tau)+\cdots+(\tau, \tau)+(\tau, \tau-1)+\cdots+(\tau, 1)$ has one of the values $0, 1, \cdots, m/d_1-1$.

The array of Fig. 1 is clearly determined by the characteristic $C(m)$.

This geometrical representation extends in an obvious manner to a space of k dimensions for the case of k variables.

* Writing $\mu^2 d_{\tau-1}$ for $[\mu(d_{\tau-1})]^2$, etc.

We illustrate the definition of a completely reduced polynomial by a very simple example.

EXAMPLE. $m = 2^2 \cdot 3$, $\mu(12) = 4$; $S(12) = \begin{pmatrix} 4 & 3 & 2 \\ 1 & 2 & 6 \end{pmatrix}$; $C(12) = \left(\begin{array}{c|c|c} 2 & 1 & 1 \\ 12 & 6 & 2 \end{array}\right)$;

$$\left(\begin{array}{cc|c|c} a_{00} & a_{10} & a_{20} & a_{30} \\ a_{01} & a_{11} & a_{21} & a_{31} \\ \hline a_{02} & a_{12} & a_{22} & a_{32} \\ \hline a_{03} & a_{13} & a_{23} & a_{33} \end{array}\right),$$

where $a_{00}, a_{10}, a_{01}, a_{11}$ may have any value $0, 1, \cdots, 11$; $a_{20}, a_{21}, a_{22}, a_{12}, a_{02}$ may have any value $0, 1, \cdots, 5$; $a_{30}, a_{31}, a_{32}, a_{33}, a_{23}, a_{13}, a_{03}$ may have either of the values $0, 1$.

For a simple numerical case the reduction of a given polynomial to a completely reduced polynomial is carried out as follows:

EXAMPLE. $f(x_1, x_2) = 29x_1^6x_2^5 + 7x_1^5 + 10x_1^2x_2$, $m = 12$.

As chain of congruences we choose $x_1(x_1-1)(x_1-2)(x_1-3) \equiv 0$, $2x_1(x_1-1)(x_1-2) \equiv 0$, $6x_1(x_1-1) \equiv 0 \pmod{12}$, and similar congruences for x_2. We find easily

$$29x_1^6x_2^5 \equiv x_1^2x_2^3 - 2x_1^2x_2 - 6x_1x_2; \quad 7x_1^5 \equiv x_1^3 + 6x_1; \quad 10x_1^2x_2 \equiv 4x_1^2x_2 + 6x_1x_2;$$

and thus $f(x_1, x_2) \equiv x_1^2x_2^3 + 2x_1^2x_2 + x_1^3 + 6x_1$, with the arrangement of coefficients

$$\left(\begin{array}{cc|c|c} 0 & 6 & 0 & 1 \\ 0 & 0 & 2 & 0 \\ \hline 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \end{array}\right)$$

which agrees with our definition of completely reduced polynomials.

4. Collecting results, we have

THEOREM IV. *For a given modulus m, and k variables $x_1, \cdots, x_k$, the signature $S(m)$ determines a chain of residual congruences* $1, \{\mu(d_i), m/d_i\}_{x_j}$. *These congruences reduce any polynomial $f(x_1, \cdots, x_k)$ to a completely reduced polynomial $g(x_1, \cdots, x_k)$. The structure of this polynomial is completely determined by the characteristic $C(m)$. The reduction is unique.*

In particular, from $S(p) = \begin{pmatrix} p \\ 1 \end{pmatrix}$ and $C(p) = \left(\left|\begin{matrix} p \\ p \end{matrix}\right|\right)$ for p a prime, it follows that for this case there is no restriction on the completely reduced polynomial beyond the obvious one that the degree of each variable is $\leq p-1$. The coefficients range independently over $0, 1, \cdots, p-1$. In all other cases we have restrictions of a very strong character.

5. It is now a simple matter to read off, for a given m, from the schedule of coefficients of the completely reduced polynomials, or from $C(m)$, the total number, $N = N(m)$, of such polynomials.

For two variables, we have $\mu^2 d_{\tau-1}$ coefficients which may assume independently each of m values, then $\mu^2 d_{\tau-1} - \mu^2 d_{\tau-2}$ coefficients which assume each $m/d_{\tau-1}$ values, etc., down to $\mu^2 d_1 - \mu^2 d_2$ coefficients each assuming m/d_2 values, and, finally, $\mu^2 m - \mu^2 d_1$ coefficients, each assuming m/d_1 values. We find, altogether (compare R. I, Theorem V),

$$N = m^{\mu^2 d_{\tau-1}} \cdot \left(\frac{m}{d_{\tau-1}}\right)^{\mu^2 d_{\tau-2} - \mu^2 d_{\tau-1}} \cdots \left(\frac{m}{d_2}\right)^{\mu^2 d_1 - \mu^2 d_2} \left(\frac{m}{d_1}\right)^{\mu^2 m - \mu^2 d_1}$$

$$= m^{\mu^2 m} \cdot d_1^{\mu^2 d_1 - \mu^2 m} \cdot d_2^{\mu^2 d_2 - \mu^2 d_1} \cdots d_{\tau-1}^{\mu^2 d_{\tau-1} - \mu^2 d_{\tau-2}}$$

$$= \left(\frac{m}{d_1}\right)^{\mu^2 m} \cdot \left(\frac{d_1}{d_2}\right)^{\mu^2 d_1} \cdots \left(\frac{d_{\tau-2}}{d_{\tau-1}}\right)^{\mu^2 d_{\tau-2}} \cdot \left(\frac{d_{\tau-1}}{1}\right)^{\mu^2 d_{\tau-1}}.$$

For polynomials of k variables, we obtain in the same manner

THEOREM V. $N = \left(\frac{m}{d_1}\right)^{\mu^k m} \cdot \left(\frac{d_1}{d_2}\right)^{\mu^k d_1} \cdots \left(\frac{d_{\tau-2}}{d_{\tau-1}}\right)^{\mu^k d_{\tau-2}} \cdot \left(\frac{d_{\tau-1}}{1}\right)^{\mu^k d_{\tau-1}}.$

The following are special cases.

COROLLARY. (a) *for $m = p$, a prime, there are no relations between the coefficients, and $N = p^{(p^k)}$;*

(b) *for $m = p_1 \cdot p_2 \cdots p_s$, $N = p_1^{p_1^k} \cdots p_s^{p_s^k}$;*

(c) *for $m = p^\lambda$, $\lambda < p$,*[*]

$$N = p^{p^k[1^k \cdot \lambda + (2^k - 1^k)(\lambda-1) + (3^k - 2^k)(\lambda-2) + \cdots + (\lambda^k - (\lambda-1)^k)\cdot 1]}.$$

RESIDUE SYSTEMS

6. We turn our attention to the individual residue systems modulo m of polynomials of more than one variable. We shall again find the rather remarkable isomorphism between the structure of the totality of the residue systems and the totality of the completely reduced polynomials which we encountered in R. I. We recall the main argument used in this connection in R. I. Defining a residue system of $f(x)$, modulo m, as the set of smallest non negative residues of $f(0)$, $f(1)$, $\cdots$, $f(m-1)$, taken in the order indicated, we saw that certain congruential relations exist between these m numbers; it was then easy to show that these relations are just sufficient to cut down the number m^m of sets which we should have if each residue assumed unrestrictedly all values $0, 1, \cdots, m-1$, to the number $N(m)$ of

[*] For the restriction $\lambda < p$, compare R. I, Theorem V.

completely reduced polynomials. From this followed that the set of all residue systems modulo m is obtained by taking into account exactly the congruential relations mentioned above.*

7. For

$$C(m) = \left(\begin{matrix}\mu(d_{\tau-1}) \\ m\end{matrix}\middle|\begin{matrix}\mu(d_{\tau-2})-\mu(d_{\tau-1}) \\ m/d_{\tau-1}\end{matrix}\middle|\cdots\middle|\begin{matrix}\mu(m)-\mu(d_1) \\ m/d_1\end{matrix}\right)$$

and one variable the residue system possesses the following structure (writing c_i for the smallest non-negative residue of $f(i)$):

$$c_0 \cdots c_{\mu(d_{\tau-1})-1} \mid c_{\mu(d_{\tau-1})} \cdots c_{\mu(d_{\tau-2})-1} \mid \cdots$$
$$\mid c_{\mu(d_2)} \cdots c_{\mu(d_1)-1} \mid c_{\mu(d_1)} \cdots c_{\mu(m)-1} \mid c_{\mu(m)} \cdots c_{m-1}.$$

The first group of $\mu(d_{\tau-1})$ residues may be chosen arbitrarily, i. e., each one may range independently over the values $0, 1, \cdots, m-1$. The second set, of $\mu(d_{\tau-2})-\mu(d_{\tau-1})$ residues, are then no longer quite unrestricted; as a result of the congruential relations between the elements of the residue system, each one of this set may assume independently any value among certain $m/d_{\tau-1}$ of the numbers $0, 1, \cdots, m-1$; the elements of the third set, of $\mu(d_{\tau-3})-\mu(d_{\tau-2})$ residues, are then more strongly restricted: each of these may assume independently any among certain $m/d_{\tau-2}$ of the numbers $0, 1, \cdots, m-1$, etc. Each of the $\mu(m)-\mu(d_1)$ elements of the last set but one may still assume independently any among certain m/d_1 of the numbers $0, 1, \cdots, m-1$, and, finally, all the $m-\mu(m)$ residues of the last set are uniquely determined. On account of the fact mentioned last, it is, in any symbolic notation for a residue system, unnecessary to consider more than the first $\mu(m)$ elements.

The congruential relations between the elements of the residue system, which determine the structure of the system, are of the following type (R. I, Theorem X):

For any modulus m and any factor d of m (including $d = m$ and $d = 1$) there exists a "*residual congruence*"

$$\frac{m}{d}\left[c_{x+\mu(d)} - \binom{\mu(d)}{1} c_{x+\mu(d)-1} + \binom{\mu(d)}{2} c_{x+\mu(d)-2} - \cdots \pm c_x\right] \equiv 0 \pmod{m},$$

which we also denote by $\}\mu(d),\ m/d\{$, or $(m/d)\cdot c_{x+\mu(d)} \equiv l \pmod{m}$, where l indicates some linear function of $c_{x+\mu(d)-1}, \cdots, c_x$, with integral coefficients in which we are not interested.

* Compare R. I, §§ 10–13.

From the signature $S(m)$ we read off a chain of residual congruences, all modulo m,

$$(m/d_i)\cdot c_{x+\mu(d_i)} \equiv l \pmod{m}, \quad \text{or} \quad \}\,\mu(d_i),\ m/d_i\,\{ \qquad (i=0,1,\cdots,\tau-1).$$

The congruences of this chain just exhaust the relations which exist between the elements of a residue system.

8. **Two variables,** $k=2$. We define our residue system to be the system of integers

$$\begin{array}{cccc} c_{00} & c_{10} & \cdots & c_{m-1,0} \\ c_{01} & c_{11} & \cdots & c_{m-1,1} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdots & \cdot \\ c_{0,m-1} & c_{1,m-1} & \cdots & c_{m-1,m-1}, \end{array}$$

where c_{ij} is the smallest non-negative residue modulo m of $f(i,j)$. We shall then have in the jth row the residues of a polynomial in x alone, $f(x,j)$, and similarly in the ith column the residues of a polynomial in y alone, $f(i,y)$. Therefore we shall certainly have in each row and each column separately the same structure as for a polynomial in one variable.

Thinking of $f(x,y)$ as a polynomial in x alone, we shall have for our residue system a structure which will be sufficiently clearly indicated by Fig. 2 (where only a first index i has been inserted instead of c_{ij}), but thinking of $f(x,y)$ as a polynomial in y alone, we shall have a structure as represented by Fig. 3 (with second index j inserted instead of c_{ij}). Treating next x, y as varying independently of each other, the two structures will be in a certain sense superimposed, and we shall have an arrangement as in Fig. 4.

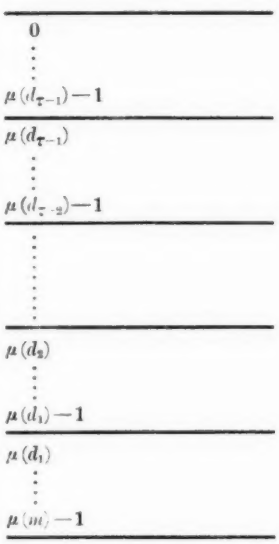


Fig. 3

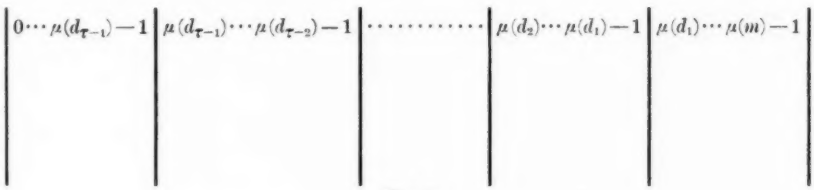


Fig. 2

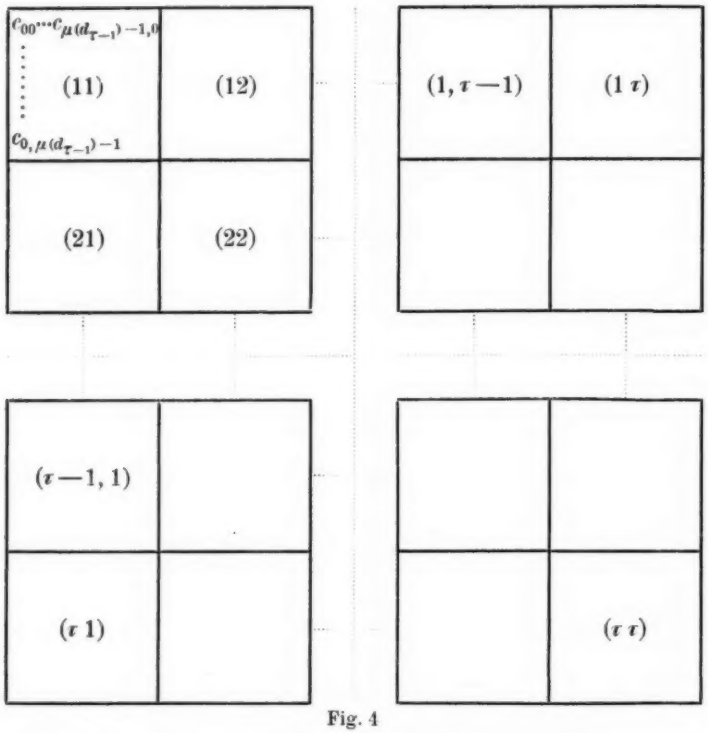


Fig. 4

In this figure we may expect the following state of affairs:

(1) all c_{ij} in (1,1), i.e., all $c_{ij}(i,j=0,\cdots,\mu(d_{\tau-1})-1)$ may unrestrictedly have any value $0, 1, \cdots, m-1$;

(2) all c_{ij} in (1, 2), (2, 2), (2, 1), i.e., all c_{ij} $(i,j=\mu(d_{\tau-1}),\cdots,\mu(d_{\tau-2})-1)$ may have any of certain $m/d_{\tau-1}$ values among $0, 1, \cdots, m-1$. The numbers from which these c_{ij} may be chosen may be expected to be determined by the choice of the c_{ij} of (1, 1);

(3) similarly for any "fringe" $(1,s)+(2,s)+\cdots+(s-1,s)+(s,s)+(s,s-1)+\cdots+(s,1)$. For the arrangement, compare Fig. 1 and text.

The system of residual congruences modulo m, as far as they are represented by our work up to the present point, will be the following:

$\mu(m)\cdot\tau$ congruences

$$\frac{m}{d_i}\left[c_{x+\mu(d_i),j}-\binom{\mu(d_i)}{1}c_{x+\mu(d_i)-1,j}+\binom{\mu(d_i)}{2}c_{x+\mu(d_i)-2,j}-\cdots\pm c_{xj}\right]\equiv 0\,(\mathrm{mod}\,m)$$

$$(i=0,1,\cdots,\tau-1;^{*}\ j=0,1,\cdots,\mu(m)-1);$$

* Since $i=\tau$ leads to a trivial congruence.

$\mu(m)\cdot\tau$ congruences

$$\frac{m}{d_j}\left[c_{i,x+\mu(d_j)}-\binom{\mu(d_j)}{1}c_{i,x+\mu(d_j)-1}+\binom{\mu(d_j)}{2}c_{i,x+\mu(d_j)-2}-\cdots\pm c_{ix}\right]\equiv 0 \pmod m$$
$$(j=0,1,\cdots,\tau-1;\ i=0,1,\cdots,\mu(m)-1).$$

We want next to show that this system of $2\mu(m)\cdot\tau$ congruences is complete in the sense that all other relations existing between the elements of a residue system are implied by it. The question of how far the $2\mu(m)\cdot\tau$ congruences are independent would be easily settled, but is not of interest for our purposes. That they are not independent is obvious.

We prove the completeness of the set of congruences in this manner (compare R. I, p. 43):

I. The number of residue systems must be $N(m)$;

II. The structure of the residue system, as determined by the chain of residual congruences above, gives exactly $N(m)$ systems (in block (1,1) there are $\mu^2 d_{\tau-1}$ numbers c_{ij}, each of which may assume independently m values; in blocks $(1,2)+(2,2)+(2,1)$ there are $\mu^2 d_{\tau-2}-\mu^2 d_{\tau-1}$ numbers, each of which has one of $m/d_{\tau-1}$ values, etc.; compare the argument of § 5).

9. **Three or more variables, $k\geq 3$.** The work is extended in an obvious manner. We omit the details, because no new type of argument is used, while the notation grows cumbersome. The final result is contained in

THEOREM VI. *The following represent a set of necessary and sufficient conditions for a system of integers, $c_{i_1\cdots i_k}$ $(i_1,i_2,\cdots,i_k=0,1,\cdots,m-1)$, $c_{i_1\cdots i_k}$ = smallest non-negative residue modulo m of $f(i_1,\cdots,i_k)$, to be a residue system modulo m of a polynomial $f(x_1,\cdots,x_k)$: Form the signature $S(m)$ and the characteristic $C(m)$; from $S(m)$ we read off k sets of $\tau\cdot\mu^{k-1}\,m^*$ residual congruences each, complete in the sense that all other residual congruences between elements of the residue system are implied by these. These sets of congruences are the following*:

$$\frac{m}{d_i}\left[c_{x+\mu(d_i),j_2,j_3,\cdots,j_k}-\binom{\mu(d_i)}{1}c_{x+\mu(d_i)-1,j_2,j_3,\cdots,j_k}+\cdots\pm c_{x,j_2,\cdots,j_k}\right]\equiv 0 \pmod m$$
$$(i=0,1,\cdots,\tau-1;\ j_2,j_3,\cdots,j_k=0,1,\cdots,\mu(m)-1);$$

$$\frac{m}{d_i}\left[c_{j_1,x+\mu(d_i),j_3,\cdots,j_k}-\binom{\mu(d_i)}{1}c_{j_1,x+\mu(d_i)-1,j_3,\cdots,j_k}+\cdots\pm c_{j_1,x,j_3,\cdots,j_k}\right]\equiv 0 \pmod m$$
$$(i=0,1,\cdots,\tau-1;\ j_1,j_3,\cdots,j_k=0,1,\cdots,\mu(m)-1);$$

. . . . . .

$$\frac{m}{d_i}\left[c_{j_1,j_2,\cdots,j_{k-1},x+\mu(d_i)}-\binom{\mu(d_i)}{1}c_{j_1,\cdots,j_{k-1},x+\mu(d_i)-1}+\cdots\pm c_{j_1,\cdots,j_{k-1},x}\right]\equiv 0 \pmod m$$
$$(i=0,1,\cdots,\tau-1;\ j_1,j_2,\cdots,j_{k-1}=0,1,\cdots,\mu(m)-1).$$

* $\mu^{k-1}m=[\mu(m)]^{k-1}$.

The structure of the residue system is determined by these congruences and can be easily written down from the characteristic $C(m)$:

(1) *the* $\mu^k d_{\tau-1}$ *numbers* $c_{i_1,\cdots,i_k}(i_1, \cdots, i_k = 0, 1, \cdots, \mu(d_{\tau-1}) - 1)$ *may assume independently any of the values* $0, 1, \cdots, m-1$;

(2) *the* $\mu^k d_{\tau-2} - \mu^k d_{\tau-1}$ *numbers* $c_{i_1,\cdots,i_k}(i_1, \cdots, i_k = \mu(d_{\tau-1}), \cdots, \mu(d_{\tau-2}) - 1)$ *may then assume independently certain* $m/d_{\tau-1}$ *values among* $0, 1, \cdots, m-1$; *for each* c *the corresponding* $m/d_{\tau-1}$ *values over which it may range are determined by the congruences above*;

(3) *the* $\mu^k d_{\tau-3} - \mu^k d_{\tau-2}$ *numbers* $c_{i_1,\cdots,i_k}(i_1, \cdots, i_k = \mu(d_{\tau-2}), \cdots, \mu(d_{\tau-3}) - 1)$ *may still assume independently* any of* $m/d_{\tau-2}$ *values among* $0, 1, \cdots, m-1$;

. . . . . .

(4) *the* $\mu^k d_2 - \mu^k d_1$ *numbers* $c_{i_1,\cdots,i_k}(i_1, \cdots, i_k = \mu(d_2), \cdots, \mu(d_1) - 1)$ *may still each assume any of* m/d_1 *values*; *finally*

(5) *the remaining* $m^k - \mu^k d_1$ *numbers* $c_{i_1,\cdots,i_k}(i_1, \cdots, i_k = \mu(d_1), \cdots, \mu(m) - 1)$ *are then completely determined among the numbers* $0, 1, \cdots, m-1$.

The number of residue systems is $N(m)$.

In saying that our theorem represents necessary and sufficient conditions for a residue system, the word *necessary* must of course be understood in the sense that it is not required to select just the intersections of the first $\mu(d_{\tau-1})$ rows and columns—for the case of two variables, and *mutatis mutandis* for $k > 2$—for the unrestricted elements, etc. For example, the intersections of any $\mu(d_{\tau-1})$ consecutive rows and any $\mu(d_{\tau-1})$ consecutive columns may be chosen as the elements which are not restricted. On the other hand, it would not do to select arbitrarily $\mu(d_{\tau-1})$ rows and $\mu(d_{\tau-1})$ columns for the unrestricted elements.

* In this sense: After the numbers c of (1) and of (2) have been chosen, the congruences will leave for each c of (3) just $m/d_{\tau-2}$ values from $0, 1, \cdots, m-1$ which it may assume. Similarly in (4) and (5).

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

A CRITERION FOR THE CONFORMAL EQUIVALENCE OF A RIEMANN SPACE TO A EUCLIDEAN SPACE*

BY

JESSE DOUGLAS

I. INTRODUCTION

The object of the present paper is to prove the following theorem.

THEOREM. *Suppose that* $\mathfrak{R}$, *a region of a Riemann space,*

$$ds^2 = \sum a_{ij}\, dx_i\, dx_j,$$

of $n \geqq 3$ *dimensions, contains a family of curves* $\mathfrak{F}$ *so distributed that a unique curve of* $\mathfrak{F}$ *passes through each pair of points of* $\mathfrak{R}$†, *and that in every triangle formed by three curves of* $\mathfrak{F}$ *the sum of the angles is two right angles. Then* $\mathfrak{R}$ *must be conformal to a euclidean space* $\mathfrak{E}$, *and in such a way that* $\mathfrak{F}$ *corresponds to the straight lines.*

As is well known, a Riemann space of more than two dimensions is, in general, not conformally equivalent to a euclidean space. The above theorem may be regarded as furnishing a criterion for such conformal equivalence.

The theorem breaks down for the case $n = 2$, not as to the conformality of the space $\mathfrak{R}$ to a space $\mathfrak{E}$, for, of course, every surface is representable conformally on a plane, but as to the possibility of turning the family $\mathfrak{F}$ conformally into the straight lines.

Consider on any surface the isogonal trajectories of an arbitrary one-parameter family of curves. The doubly infinite totality of these trajectories

* Presented, under another title, to the Society, October 27, 1923. The results of this paper include those of a previous paper by the author, *Determination of all systems of* ∞^4 *curves in space in which the sum of the angles of every triangle is two right angles,* Bulletin of the American Mathematical Society, vol. 29 (1923), pp. 356–366.

† By taking the two points infinitely close to one another, this implies that there is a unique curve of $\mathfrak{F}$ through each point in each direction. The number of curves in $\mathfrak{F}$ is necessarily ∞^{2n-2}. The most usual way of defining a curve system of the type of $\mathfrak{F}$ is by a set of differential equations of the form

$$\frac{d^2x_i}{ds^2} = F_i\left(x_j, \frac{dx_j}{ds}\right).$$

We assume whatever degree of regularity in the functions F_i is necessary to assure the function-theoretic validity of the arguments that follow.

is termed an *isogonal* family. It is easily proved synthetically that in every isogonal family the sum of the angles of any triangle formed by three of the curves is two right angles, and, conversely, that this property is characteristic.* But the isogonal family is conformally equivalent to the straight lines of a plane only when the base family of ∞^1 curves is isothermal.†

II. Equivalence of the family $\mathfrak{F}$ to the straight lines by point transformation

Our method of proof will be entirely synthetic. For definiteness, and in order to have concrete figures to deal with, we take $n = 3$, but our argument is easily extended to higher values of n.

First, we show that under the hypothesis of the two-right-angle property the family $\mathfrak{F}$ must be point-wise equivalent to the straight lines of a projective space.

The situation here is governed by the following fundamental theorem of projective geometry.‡

Suppose that in a region of space we have a family of curves C and a family of surfaces Σ arranged as follows:

1. There is a unique curve C through any given two points and a unique surface Σ through any given three points of the region, provided that the three points are not on the same curve C.

2. The curve C which joins any two points of a surface lies wholly on Σ.

Then it is possible to introduce coördinates into the space so that the curves C and the surfaces Σ are represented by linear equations.

In other words, the curves C and surfaces Σ can by point transformation be converted into the straight lines and planes of a (region of a) projective space.

Our problem is, of course, to prove that the possession of the two-right-angle property by the family $\mathfrak{F}$ implies the conditions of the theorem just stated.

Each point together with a direction through it determines a curve of $\mathfrak{F}$. Consider the curves of $\mathfrak{F}$ which pass through any fixed point O in the ∞^1 directions of any linear pencil. They form a surface Σ which is easily shown to have the property that the curve C of $\mathfrak{F}$ which joins any two points of Σ lies wholly on Σ.

* G. Scheffers, *Isogonalkurven, Äquitangentialkurven und komplexe Zahlen*, Mathematische Annalen, vol. 60 (1905), p. 504.

† E. Kasner, Mathematische Annalen, 1904.

‡ See Felix Klein, *Über einen Satz aus der Analysis Situs, Gesammelte Mathematische Abhandlungen*, erster Band, pp. 306-310; also *Über die sogenannte nicht-euklidische Geometrie* in the same volume, pp. 311-343, especially pp. 333-343.

For let P and Q be any two points of Σ. These must lie respectively on two curves C_1 and C_2 of $\mathfrak{F}$ which pass through O in directions d_1 and d_2 belonging to the pencil defining Σ. Let R be any point on the curve C of $\mathfrak{F}$ which joins P and Q, and let the curve OR have at O the direction d. Then, by applying the hypothesis of an angle sum of two right angles to the triangles OPR, OQR, OPQ, we can deduce that

$$\sphericalangle\, d_1 d + \sphericalangle\, d d_2 = \sphericalangle\, d_1 d_2.$$

This relation implies that d belongs to the pencil of d_1 and d_2; therefore the curve OR is one of those composing the surface Σ; hence the point R lies on Σ. Thus the curve C lies entirely on Σ, since R was an arbitrary point of C.

It remains only to show that there is one and only one surface Σ through any given three points not on the same curve of $\mathfrak{F}$. As defined above, a surface Σ appears with a special point O marked upon it, but this point O is not really distinguished from any other point on the surface. More precisely, if O' is any point of Σ, then Σ is identical with the locus Σ' of the curves C of $\mathfrak{F}$ which pass through O' in the pencil of directions tangential to Σ at O'.

First, if P is any point of Σ, construct the curve C of $\mathfrak{F}$ which joins O' and P. This curve lies entirely on Σ; therefore it has at O' a direction tangential to Σ; therefore C belongs to Σ'; therefore P belongs to Σ'.

Second, if P' is any point of Σ', it lies on a curve C' which has at O' a direction d' tangential to Σ. Allow a point Q which remains always on Σ to approach O' in the direction d'. Then the curve $O'Q$ of $\mathfrak{F}$ always lies on Σ, and evidently approaches to C' as its limit. Hence C' and therefore P' lies on Σ.

Having established the identity of Σ and Σ', let us consider any three points P, Q, R of space not on the same curve of $\mathfrak{F}$. Then the curves PQ, PR have at P two distinct directions d_1, d_2, which determine a pencil. Any surface Σ containing P, Q, R must contain P and have the directions of the pencil $d_1 d_2$ for tangential directions at P; these conditions, by the above, make Σ unique.

With this we have completed our proof of the existence of a point transformation throwing $\mathfrak{F}$ in the Riemann space $\mathfrak{R}$ into the straight lines $\mathfrak{S}$ of a projective space $\mathfrak{P}$.

III. The cones of the angular metric

The measurement of angles in $\mathfrak{R}$ is defined by the (elemental) quadric cones

K: $$\sum a_{ij}\, dx_i\, dx_j = 0,$$

one of which is associated with each point O of the Riemann space. In fact, the angle between any two directions at O is $1/2i$ times the logarithm of the anharmonic ratio of these two directions relative to the two cut from the cone associated with O by the plane determined by the first two.

The projective space $\mathfrak{P}$ is metrically amorphous, having, in particular, no metric for the measurement of angles. Suppose we endow it with such a metric by allowing the point transformation which converts $\mathfrak{F}$ into the straight lines to act also on the cones K of $\mathfrak{R}$. Then each of these goes over into a cone of elements through a point of the space $\mathfrak{P}$, and if we imagine each element extended into a complete straight line, we have a complete quadric cone associated with each point of $\mathfrak{P}$. We then define angles in $\mathfrak{P}$, as in $\mathfrak{R}$, by the logarithm of an anharmonic ratio.

We denote by $\mathfrak{P}'$ the space $\mathfrak{P}$ as provided with its angular metric.* Since every point transformation is linear im kleinen, and linear transformations preserve anharmonic ratios, the transformation from $\mathfrak{R}$ to $\mathfrak{P}'$ is angle conserving. And since the transformation throws $\mathfrak{F}$ into the straight lines of $\mathfrak{P}'$, the sum of the angles of any rectilinear triangle in $\mathfrak{P}$ must be two right angles.

The purpose of the next two sections is to show that this implies that the cones K of $\mathfrak{P}'$ pass through a fixed conic I.

In order to avoid interrupting the main thread of our argument to dispose of possible exceptions analogous to the minimal lines and planes of ordinary space, we presume the quadratic form of $\mathfrak{R}$ to be positive definite, so that the cones K are nullpartite (have no real points except the vertex). This restriction is quite unessential, our conclusion being just as valid for indefinite quadratic forms.

IV. The case of the plane

We have a configuration consisting of the straight lines $\mathfrak{S}$ and the cones K of the space $\mathfrak{P}'$. Imagine that we take a section of this configuration by an arbitrary plane π. Then we obtain in π, first, its straight lines, and, second, associated with point P of π a pair of (conjugate imaginary) lines through P by means of which angles at P are to be measured.

Our goal in this section is to prove that the two-right-angle condition for the rectilinear triangle of π implies that the two lines associated with an arbitrary point of π pass respectively through two fixed points I and J ("circular points at infinity").

In doing this the following lemma will be very useful.

* Distance is not defined in $\mathfrak{P}'$.

LEMMA. *If each of three points A, B, C sends its lines through two fixed points I and J, respectively, then the same must be true of an arbitrary point of their plane.*

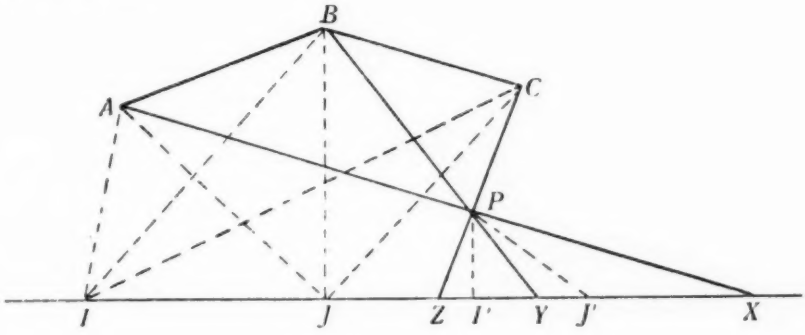


Fig. 1

Proof. Let P be an arbitrary fourth point in the plane of A, B, C, and let the line pair of P intersect the (real) line λ of I, J in I', J'; also let AP, BP, CP intersect λ in X, Y, Z respectively. Then it is easy to show that the condition of an angle sum of two right angles for the triangle ABP expresses itself as follows in terms of anharmonic ratios:

$$(XYIJ) = (XYI'J').$$

The triangle BCP gives

$$(YZIJ) = (YZI'J'),$$

and from this and the preceding follows

$$(XZIJ) = (XZI'J'),$$

which can also be obtained directly from the triangle ACP. These three relations imply that the point pair IJ is identical with $I'J'$, which is what was to be proved.

Our argument is not at all invalidated if A, B, C are collinear. The above proof still holds for every point P not on the line ABC, and we can then extend it to any point of this line by using any three of the points P (or by continuity).

This lemma established, the next step is to cause to correspond to each point pair AB a certain (real) line λ, as follows. The line pairs of A and B form a complete quadrilateral having two diagonals besides AB (both real).

Suppose we select one of these for a particular point pair; then the selection is determined by continuity for every point pair, at least within a certain region of the continuum of point pairs.

Now there are ∞^4 point pairs and only ∞^2 straight lines in the plane π. Therefore there are just these three possibilities: (1) the ∞^4 point pairs correspond in sets of ∞^2 to ∞^2 straight lines, or (2) in sets of ∞^3 to ∞^1 straight lines, or (3) all the ∞^4 correspond to one line.* In any case there certainly exists a line λ together with ∞^2 point pairs AB whose associated line pairs intersect in two points I, J of λ.

Let AB and A_1B_1 be any two of the ∞^2 point pairs associated with λ, and let A, B send their line pairs through the points I, J of λ while A_1, B_1 send theirs through I_1, J_1 of λ. Extend AA_1, AB_1, BA_1, BB_1 to meet λ in X, Y, Z, W respectively.

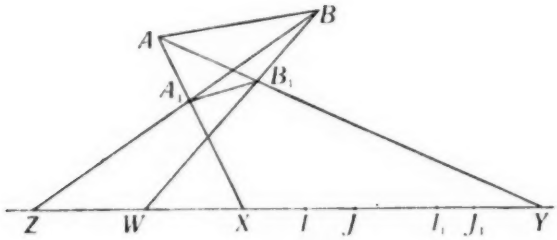


Fig. 2

Applying the two-right-angle condition to the triangles ABA_1, ABB_1, AA_1B_1, BA_1B_1, we find that the point pairs IJ and I_1J_1 must have equal anharmonic ratios with respect to any two of the four points X, Y, Z, W. The consequence of this is that I_1J_1 is identical with IJ. We thus have four points sending their lines through I and J, which by our lemma is enough to insure the same condition for every point of the plane.

The above reasoning breaks down only when the line of A_1B_1 is identical with the line of AB, in which case the four points X, Y, Z, W reduce to one point.†

But if this is true for every two of the ∞^2 point pairs associated with λ, these ∞^2 point pairs must be the totality of point pairs of a fixed line μ, or some two-dimensional portion of that totality. Then there must be two of them such as AB and AC, having a point A in common, so that the three points A, B, C have their lines going through the same two points I, J

* A finite number different from one is ruled out by the above lemma, also by continuity.

† It does not break down in a case such as where the line AB contains A_1, when the four points X, Y, Z, W reduce to three.

of λ, from which the result we wish to establish follows no less than before.

V. The conic Γ

Consider any line l of space. We begin in this section by proving that the cones associated with the different points of l all pass through the same conic Γ_l.*

Any plane π through l cuts from the ∞^1 cones related to l ∞^1 (conjugate imaginary) line pairs, all of which pass through two fixed points I, J in π. Furthermore, I, J are the only points common to all the line pairs and therefore the only points in π common to all the cones.

Imagine π to rotate about l. Then the points I, J describe a certain locus, which is common to all the cones of l. Hence these ∞^1 cones all intersect in a common algebraic curve, which must be a conic, since it is cut in precisely two points by an arbitrary plane through l. It must be furthermore a proper conic, for the only alternative is that the cones of l have a pair of generators in common, an eventuality easily disposed of.

An alternative proof is to say that the common curve of the ∞^1 quadric cones cannot be a skew quartic, for only four quadric cones can pass through a skew quartic; it cannot be a skew cubic, for such a curve is projected only from its own points by quadric cones, whereas l cannot be part of the skew cubic; therefore the common curve must be a conic.

This association of a conic with each line l of space produces the association of a (real) plane π_l, the plane of the conic, with each line l. π_l contains the "line at infinity" IJ of every plane through l. π_l does not contain l, since every plane through l has only two points and not a conic in common with all the cones of l.

Now the planes π_l must in fact reduce to a single plane. For the proof, let O be any point of space whose cone does not degenerate into a plane. Suppose l and m to be any two lines through O, and n any third line through O not coplanar with the first two. Each of the planes π_l, π_m, π_n intersects, since it does not contain, its line in a unique point; let the three points of intersection be respectively A, B, C. Then A, B, C are three distinct points. Otherwise two of them would fall together in O; this would cause the plane π_l, for instance, to contain O. It is easily seen that the cone of O would then degenerate into the plane π_l, contrary to assumption.

* The only assumption made relative to l is that its cones shall be proper cones, not degenerating into planes, except perhaps for isolated points on l. We assume, of course, that the cones do not degenerate all over space (quadratic form of $\mathfrak{R}$ of rank n). Then the locus of points whose cones degenerate is at most a surface, analogous to the plane at infinity in ordinary space, and we suppose l not to lie on this surface.

Each of the planes π_l, π_m intersects the plane lm in the "line at infinity" of lm. Therefore this line must be the line AB since π_l and lm contain A and π_m and lm contain B. Thus π_l contains B, and by a similar argument it contains C. The same reasoning applies to π_m, so that π_l and π_m must fall together in the plane ABC.

On account of the arbitrary choice of l and m, it follows that all the lines through O determine the same plane Π. This means that the cones of any two points of a line through O intersect in a conic lying in Π. Hence if Γ is the conic in which the cone of O intersects Π, the cone of every point of space must pass through Γ.

VI. Representation on a euclidean space

It is evident how to bring our argument to its conclusion. We have only to introduce a euclidean space $\mathfrak{E}$ and to collineate the projective angularly-metric space $\mathfrak{P}'$ into $\mathfrak{E}$ so that the conic Γ in $\mathfrak{P}'$ goes over into the absolute circle (Kugelkreis) of $\mathfrak{E}$. Obviously each angle of $\mathfrak{P}'$ is by this transformation equal to the corresponding angle of $\mathfrak{E}$.

By combining the representation of $\mathfrak{R}$ on $\mathfrak{P}'$ with that of $\mathfrak{P}'$ on $\mathfrak{E}$, we have a transformation of $\mathfrak{R}$ into $\mathfrak{E}$ which conserves angles and converts the family $\mathfrak{F}$ into the straight lines.

Columbia University,
New York, N.Y.

THE DEFLECTION OF A RECTANGULAR PLATE FIXED AT THE EDGES*

BY

H. W. MARCH

The problem of finding the deflection of a loaded rectangular plate fixed at the edges has offered considerable analytical difficulty. A very ingenious method of approximation capable of application to many other boundary value problems was applied to this problem by Ritz† who stated that the usual method of combining particular solutions of the appropriate differential equation is known to fail. Presumably his reason for this statement was that the latter method leads to an infinite system of equations in infinitely many variables whose solution presents difficulty. In the present treatment the method of combining particular solutions is used and the unique solution of the resulting infinite system of equations for the determination of the coefficients is given. The deflection is expressed in a form that appears to be simpler and better adapted to the discussion of the elastic surface than that given by Ritz. In the case of a uniformly loaded plate to which the discussion of this paper is confined, the deflection is found to a high degree of approximation by retaining the first term of the infinite series in the expression of the solution.

A treatment of the same subject, making use of particular solutions of the type used in this paper, was given, as I have recently discovered, by H. Hencky in a Darmstadt dissertation in 1913. He failed, however, to give an adequate discussion of the infinite system of equations to which he was led. In the present paper a fortunate choice of coefficients in the infinite series representing the deflection results in a much more tractable system of equations than Hencky's, the unknown coefficients decreasing much more rapidly in numerical value with increasing index. There results a very decided shortening of the numerical work. For example, Hencky finds for the first few coefficients in the case of a square plate

$$1.4138, \quad 0.0945, \quad -0.1093, \quad 0.0770, \quad -0.0510, \quad 0.0376.$$

The corresponding coefficients in this paper are

$$0.167, \quad -0.0010, \quad -0.00044, \quad -0.00015, \quad -0.000026, \quad -0.000015.$$

* Presented to the Society, December 29, 1923.

† W. Ritz, Journal für Mathematik, vol. 135 (1908), pp. 1–61. Also *Gesammelte Werke.*

Hencky's work has apparently escaped the attention of those interested in the theoretical aspects of elasticity.*

The discussion will be confined to the uniformly loaded plate. The method employed can probably be readily extended to the case of a load symmetrically distributed with respect to the center of the rectangle and probably without great difficulty to the case of an arbitrary distribution of load.

The solution in series. The deflection w of a uniformly loaded plate satisfies the differential equation†

$$\nabla^4 w = A, \tag{1}$$

where‡

$$A = \frac{p}{\frac{2}{3}Eh^3/(1-\sigma^2)},$$

p being the uniform load, h one-half the thickness of the plate, E and σ elastic constants of the material of the plate.

For a rectangular plate clamped at the edges w is subject to the boundary conditions

$$w = 0, \quad x = 0, \quad x = a, \qquad y = 0, \quad y = b; \tag{2,a}$$

$$\frac{\partial w}{\partial x} = 0, \quad x = 0, \quad x = a. \qquad \frac{\partial w}{\partial y} = 0, \quad y = 0, \quad y = b. \tag{2,b}$$

Under these conditions (1) has not more than one§ solution. To obtain the solution, let

$$w_1 = \alpha x(x-a)\,y(y-b), \tag{3}$$

where

$$\alpha = \frac{A}{8}.$$

w_1 satisfies the differential equation (1) and the boundary conditions (2,a) but does not satisfy the conditions (2,b), since

$$\begin{aligned} \frac{\partial w_1}{\partial x} &= -\alpha a y(y-b), \quad x = 0, \qquad \frac{\partial w_1}{\partial y} = -\alpha b x(x-a), \quad y = 0, \\ &= \alpha a y(y-b), \qquad x = a; \qquad\qquad = \alpha b x(x-a), \qquad y = b. \end{aligned} \tag{4}$$

To satisfy the remaining boundary conditions (2,b) add to w_1 a properly chosen solution of

$$\nabla^4 w = 0. \tag{5}$$

* Love, *Theory of Elasticity*, 3d edition, 1920, p. 496.

† Love, p. 489.

‡ Love, p. 470 and p. 489.

§ E. Mathieu, Journal de Mathématiques, vol. 14 (1869), p. 392.

If $\xi(x, y)$ is a solution of $\nabla^2 \xi = 0$ it is readily found that

$$\psi = u(x, y)\, \xi(x, y) \tag{6}$$

is a solution of (5) if u is of the form

$$u = h(x^2 + y^2) + kx + ly + m.$$

Choose the following solution of (5) which is built up from solutions of the type (6):

$$\psi_2 = \alpha \Big\{ \sum_n \frac{a_n}{\lambda_n b + \sinh \lambda_n b} [(y-b)\sinh \lambda_n y + y \sinh \lambda_n (y-b)] \sin \lambda_n x + \sum_m \frac{b_m}{\sigma_m a + \sinh \sigma_m a} [(x-a)\sinh \sigma_m x + x \sinh \sigma_m (x-a)] \sin \sigma_m y \Big\} \tag{7}$$

where

$$\lambda_n = \frac{n\pi}{a}, \qquad \sigma_m = \frac{m\pi}{b}. \tag{8}$$

Since the solution ψ_1 and the boundary conditions (2, b) are symmetrical with respect to the center of the rectangle, it is evident that only odd indices n and m will occur in the series in (7). Hereafter $\sum_m$ and $\sum_n$ will be understood to refer to summations taken over the odd integers from 1 to ∞ unless otherwise noted.

Write

$$\psi = \psi_1 + \psi_2, \tag{9}$$

and note that ψ satisfies the conditions (2, a). The coefficients a_n and b_m are to be so determined that the conditions (2, b) are satisfied:

$$\left(\frac{\partial \psi}{\partial x}\right)_{x=0} = \alpha \Big\{ -ay(y-b) + \sum_n \frac{\lambda_n a_n}{\lambda_n b + \sinh \lambda_n b} [(y-b)\sinh \lambda_n y + y \sinh \lambda_n (y-b)] - \sum_m b_m \sin \sigma_m y \Big\} = 0, \tag{10}$$

$$\left(\frac{\partial \psi}{\partial y}\right)_{y=0} = \alpha \Big\{ -bx(x-a) - \sum_n a_n \sin \lambda_n x + \sum_m \frac{\sigma_m b_m}{\sigma_m a + \sinh \sigma_m a} [(x-a)\sinh \sigma_m x + x \sinh \sigma_m (x-a)] \Big\} = 0. \tag{11}$$

If these conditions are satisfied, it is evident from the symmetry with respect to the center of the rectangle that the conditions on the edges $x = a$ and $y = b$ are also satisfied.

Equation (10) is satisfied if

$$b_m = \frac{2}{b}\int_0^b \Big\{ -ay(y-b) + \sum_n \frac{\lambda_n a_n}{\lambda_n b + \sinh \lambda_n b}\left[(y-b)\sinh \lambda_n y + y \sinh \lambda_n (y-b)\right] \sin \sigma_m y \Big\} dy. \tag{10, a}$$

The inversion of the order of summation with respect to m and n that is here implied is seen to be justified by the values found for a_n and b_m. For it is found that $|a_n| < H/(\lambda_n^{5/2})$, H a fixed positive number, and b_m satisfies a similar condition. If the second term in (10) be written $\sum_n \sum_m A_{n,m} \sin \sigma_m y$ the coefficients $A_{n,m}$ then satisfy the condition

$$|A_{n,m}| < \frac{D\sigma_m}{(\lambda_n^2 + \sigma_m^2)^2}, \quad D \text{ a fixed positive number.}$$

Hence the order of summation can be interchanged.

It follows readily from (10, a) that

$$b_m = \frac{8a}{b\sigma_m^3} - \frac{8\sigma_m}{b}\sum_n \frac{\lambda_n^2}{(\lambda_n^2 + \sigma_m^2)^2}\varrho(\lambda_n b)\, a_n, \tag{12}$$

where

$$\varrho(\lambda_n b) = \frac{1 + \cosh \lambda_n b}{\lambda_n b + \sinh \lambda_n b}. \tag{13}$$

In like manner from (11)

$$a_n = \frac{8b}{a\lambda_n^3} - \frac{8\lambda_n}{a}\sum_m \frac{\sigma_m^2}{(\lambda_n^2 + \sigma_m^2)^2}\varrho(\sigma_m a)\, b_m. \tag{14}$$

Writing

$$b = \eta a \tag{15}$$

where η is taken to be less than or equal to 1 and interchanging m and n in (12), (12) and (14) become

$$b_n = \frac{8\eta^2 a^3}{n^3\pi^3} - \frac{8n\eta^2}{\pi}\sum_m \frac{m^2}{(m^2\eta^2 + n^2)^2}\varrho(m\pi\eta)\, a_m, \tag{16}$$

$$a_n = \frac{8\eta a^3}{n^3\pi^3} - \frac{8n\eta^2}{\pi}\sum_m \frac{m^2}{(m^2 + n^2\eta^2)^2}\varrho\left(\frac{m\pi}{\eta}\right) b_m. \tag{17}$$

The function ϱ occurring in these equations does not differ much from unity for most values of its argument and its presence does not materially complicate the numerical work of solving these equations.

The square plate. The simplest case is that of the square $b = a$, i. e., $\eta = 1$. In this case $a_n = b_n$ and (16) and (17) become identical, viz.,

$$b_n = \frac{8a^3}{n^3\pi^3} - \frac{8n}{\pi}\sum_m \frac{m^2}{(m^2+n^2)^2}\varrho(m\pi)b_m, \tag{18}$$

or

$$b_n = \frac{K}{n^3} - \sum_m \alpha_{nm} b_m \qquad (n = 1, 3, 5, \cdots), \tag{19}$$

where

$$K = \frac{8a^3}{\pi^3}, \qquad \alpha_{nm} = \frac{8}{\pi}\frac{nm^2}{(n^2+m^2)^2}\varrho(m\pi). \tag{20}$$

Equations (19) can be written

$$\alpha_{n1}b_1 + \alpha_{n3}b_3 + \cdots + (1+\alpha_{nn})b_n + \cdots = \frac{K}{n^3} \quad (n = 1, 3, 5, \cdots). \tag{21}$$

The coefficient of the term in the principal diagonal rapidly approaches 1 as a limit. Before discussing the solution of the infinite system of equations (21) the first four equations of the system will be written down:

$$\begin{aligned}
&1.546\ b_1 + 0.229 b_3 + 0.094 b_5 + 0.050 b_7 + \cdots = K,\\
&0.065\ b_1 + 1.212 b_3 + 0.165 b_5 + 0.111 b_7 + \cdots = 0.03705 K,\\
&0.016\ b_1 + 0.099 b_3 + 1.128 b_5 + 0.114 b_7 + \cdots = 0.008 K,\\
&0.0061 b_1 + 0.048 b_3 + 0.081 b_5 + 1.091 b_7 + \cdots = 0.00292 K.
\end{aligned} \tag{22}$$

It will now be shown by a slight modification of an argument due to von Koch* that the system (21) has one and only one solution. For the sake of easy reference to von Koch's paper, replace the unknowns $b_1, b_3, \cdots$ by x_1, x_3 etc., and write (21) as

$$\alpha_{i1}x_1 + \alpha_{i3}x_3 + \cdots + (1+\alpha_{ii})x_i + \cdots = \frac{K}{i^3} \qquad (i = 1, 3, 5, \cdots) \tag{23}$$

where

$$\alpha_{ik} = \frac{8}{\pi}\frac{ik^2}{(i^2+k^2)^2}\frac{1+\cosh k\pi}{k\pi+\sinh k\pi} < \frac{8}{\pi}\frac{ik^2}{(i^2+k^2)^2}.$$

Divide (23) by $(1+\alpha_{ii})$ and set

$$\frac{\alpha_{ik}}{1+\alpha_{ii}} = -b_{ik}, \qquad b_{ii} = 0, \qquad \frac{K}{i^3(1+\alpha_{ii})} = c_i.$$

* H. von Koch, Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 22 (1913), pp. 285-291.

Then

$$|b_{ik}| < \frac{8}{\pi} \frac{ik^2}{(i^2+k^2)^2}, \qquad C_i < \frac{K}{i^3}.$$

Equations (23) become

$$x_i - \sum_k b_{ik} x_k = C_i \qquad (i = 1, 3, 5, \ldots). \tag{24}$$

Let

$$x_k = x_k^{(1)} + C_k.$$

Substitute and obtain

$$x_i^{(1)} - \sum_k b_{ik} x_k^{(1)} = \sum_k b_{ik} C_k = C_i^{(1)}.$$

Let

$$x_k^{(1)} = x_k^{(2)} + C_k^{(1)},$$

and so on, continuing this process of successive approximation indefinitely.

Using the notation

$$b_{ik}^{(2)} = \sum_v b_{iv} b_{vk}, \qquad b_{ik}^{(3)} = \sum_v b_{iv}^{(2)} b_{vk}, \text{ etc.},$$

it will be shown that the series

$$x_i = C_i + C_i^{(1)} + C_i^{(2)} + \cdots,$$

i. e.

$$x_i = C_i + \sum_k b_{ik} C_k + \sum_k b_{ik}^{(2)} C_k + \sum_k b_{ik}^{(3)} C_k + \cdots, \tag{25}$$

is absolutely convergent and that it represents the unique solution of the system (24), subject to the condition $|x_k| < X$, a fixed number. Von Koch's argument assumes that $\sum_k |b_{ik}| < \epsilon < 1$, ϵ the same for all values of i. This condition is not satisfied in this case, although it is true that $\sum_k |b_{ik}| < 1$. Accordingly his argument has been modified, taking advantage of the fact that in the present case $C_k < K/k^3$ instead of being simply less than a fixed number. This modification results in showing the dependence of x_i on i which is important for the discussion of the derivation of the system of equations (12) and (14). We have

$$\sum_k |b_{ik} C_k| < \frac{8}{\pi} K \sum_k \frac{ik^2}{(i^2+k^2)^2} \frac{1}{k^3} < \frac{8Ki}{\pi} \sum_k \frac{1}{\sqrt{k}\,(i^2+k^2)^2}.$$

Noting that the summation is over odd integers and taking account of the situation at the lower limit, it is readily found that

$$\sum_k \frac{1}{\sqrt{k}\,(i^2+k^2)^2} < \frac{1}{2} \int_0^\infty \frac{dk}{\sqrt{k}\,(i^2+k^2)^2}$$

$$= \frac{1}{2\,i^{7/2}} \frac{\Gamma(\frac{1}{4})\,\Gamma(\frac{7}{4})}{2\,\Gamma(2)} = \frac{1}{i^{7/2}} \frac{3}{8\sqrt{2}}.$$

Hence

$$\sum_k |b_{ik} C_k| < \frac{3}{\pi\sqrt{2}} \frac{K}{i^{5/2}};$$

$$\sum_k |b_{ik}^{(2)} C_k| \leqq \sum_k \sum_v |b_{iv}| \, |b_{vk}| \, C_k$$

$$< \frac{3K}{\pi\sqrt{2}} \sum_v \frac{|b_{iv}|}{v^{5/2}} < \frac{3K}{\pi\sqrt{2}} \frac{8i}{\pi} \sum_v \frac{1}{(i^2+v^2)^{\frac{3}{2}}\sqrt{v}}$$

$$< \left(\frac{3}{\pi\sqrt{2}}\right)^2 \frac{K}{i^{5/2}}.$$

In general

$$\sum_k |b_{ik}^{(p)} C_k| < \left(\frac{3}{\pi\sqrt{2}}\right)^p \frac{K}{i^{5/2}}.$$

Hence

$$|x_i| < \frac{K}{i^{5/2}} (1 + \varepsilon + \varepsilon^2 + \cdots) = \frac{K}{i^{5/2}} \frac{1}{1-\varepsilon}, \tag{26}$$

where

$$\varepsilon = \frac{3}{\pi\sqrt{2}} = 0.677.$$

Hence the absolute convergence of (25) is established. The proof that (25) is a solution of (24) now follows immediately, exactly as in von Koch's article. That it is the only solution subject to the restriction $|x_k| < X$, a fixed number, follows also from the method of von Koch, noting that with

$$A_i = \frac{8i}{\pi}, \qquad \varepsilon = \frac{3}{\pi\sqrt{2}}, \quad \text{and} \quad H = \sum_k \frac{1}{k^{3/2}}$$

it follows that

$$\sum_k |b_{ik}| < 1, \qquad |b_{ik}^{(2)}| < A_i \varepsilon \frac{1}{k^{3/2}}, \qquad \sum_k |b_{ik}^{(2)}| < A_i \varepsilon H,$$

$$\sum_k |b_{ik}^{(3)}| < A_i \varepsilon^2 H \quad \text{and} \quad \sum_k |b_{ik}^{(p)}| < A_i \varepsilon^{p-1} H.$$

It is to be observed that these inequalities would have been sufficient to show the absolute convergence of (25) but would not have shown the important fact that $|x_i| < M/i^{5/2}$ where M is a fixed number.

Solving the system (21) using the formula (25) it is found that

$$b_1 = 0.6477\, K, \qquad b_3 = -0.0040\, K.$$

These values are found more easily by solving the first equation for b_1, retaining only the first term, using this value in the second equation to

obtain an approximate value of b_3 by retaining the first two terms only, then correcting the value of b_1 by substituting the value found for b_3 in the first equation and so on. By this method it is found that

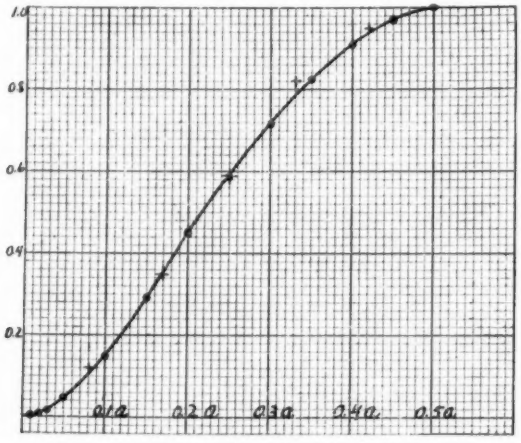
$$b_1 = 0.6475\,K, \qquad b_3 = -0.0040\,K,$$

in agreement with the foregoing results. By the same method it is also found that $b_5 = -0.0017\,K$, $b_7 = -0.006\,K$, $b_9 = -0.0003\,K$, $b_{11} = -0.0001\,K$, $b_{13} = -0.00006\,K$. This method of successive approximation though lacking complete justification can be applied with greater confidence when it is known that a solution exists and when it is known that the unknowns b_i decrease very rapidly in absolute value, being subject as shown to the relation

$$b_i < \frac{M}{i^{5/2}}.$$

Consequently the infinite series constituting the left-hand members of (21) converge very rapidly.

The deflection at any point of a square plate is given by (9) where ${}^1\Psi_1$ and ${}^1\Psi_2$ are defined by (3) and (7) respectively. It is easy to see from the values just found for the coefficients that it is necessary to retain in ${}^1\Psi_2$ only the terms containing $a_1 = b_1$.



Deflection along a diameter of a square plate

In the accompanying figure, the deflection along a median line, $x = 0.5a$, of a square plate is shown in terms of the deflection at the center. The circles indicate points determined by the present formula. The crosses

indicate results for one of the square plates experimentally investigated by Crawford.* The results for this and other plates were tabulated by Knott† for comparison with the results of the application of Ritz's method.

The rectangular plate. In (16) and (17) replace b_m by x_{m+1} and a_m by x_m, m being odd in both cases, so that an even subscript correspond to a b, an odd subscript to an a. The equations (16) and (17) can then be written

$$
\begin{aligned}
&x_1+a_{12}x_2 \quad +a_{14}x_4 \quad +a_{16}x_6 \quad +a_{18}x_8+\cdots = K = C_1,\\
&a_{21}x_1+ \quad x_2+a_{23}x_3 \quad +a_{25}x_5 \quad +a_{27}x_7 \quad = K_1 = C_2,\\
&a_{32}x_2+ \quad x_3+a_{34}x_4 \quad +a_{36}x_6 \quad +a_{38}x_8 \quad = \frac{K}{27} = C_3,\\
&a_{41}x_1 \quad +a_{43}x_3+ \quad x_4+a_{45}x_5 \quad +a_{47}x_7 \quad = \frac{K_1}{27} = C_4,\\
&a_{52}x_2 \quad +a_{54}x_4+ \quad x_5+a_{56}x_6 \quad +a_{58}x_8 \quad = \frac{K}{125} = C_5,\\
&a_{61}x_1 \quad +a_{63}x_3 \quad +a_{65}x_5+ \quad x_6+a_{67}x_7 \quad = \frac{K_1}{125} = C_6,
\end{aligned}
\tag{27}
$$

where

$$
\begin{aligned}
a_{nm} &= \frac{8\eta^2}{\pi}\,\frac{n(m-1)^2}{[n^2\eta^2+(m-1)^2]^2}\,\varrho\left(\frac{(m-1)\pi}{\eta}\right), && n \text{ odd, } m \text{ even},\\
&= \frac{8\eta^2}{\pi}\,\frac{(n-1)m^2}{[(n-1)^2+m^2\eta^2]^2}\,\varrho(m\pi\eta), && n \text{ even, } m \text{ odd},\\
&= 0, && n \text{ and } m \text{ both even or both odd and } n \neq m,\\
&= 1 && n = m,
\end{aligned}
$$

and

$$K = \frac{8\eta a^3}{\pi^3}, \qquad K_1 = \frac{8\eta^2 a^3}{\pi^3},$$

and

$$
\begin{aligned}
C_n &= \frac{K}{n^3}, && n \text{ odd},\\
&= \frac{K_1}{n^3}, && n \text{ even},
\end{aligned}
$$

and

$$K_1 < K.$$

* W. T. Crawford, Proceedings of the Royal Society of Edinburgh, vol. 32 (1911-12), pp. 348-389.

† C. G. Knott, Proceedings of the Royal Society of Edinburgh, vol. 32 (1911-12), pp. 390-392.

Adopting the notation used in the preceding section, let

$$b_{ik} = -a_{ik}, \qquad i \neq k \quad \text{and} \quad b_{ii} = 0$$

and use the method of successive approximation used in solving (24). In the following discussion $\sum_{k_e}\sum_{k_o}$ will indicate summation over even integers and odd integers respectively. When no subscript is written, the reader can readily determine which type of summation is meant. It is found that if i is odd and η so large that $\varrho(\pi\eta)<1$,

$$\sum_{k_e}|b_{ik}C_k| < \frac{\eta^2}{\pi}K\sum_{k_o}\frac{ik^2}{(i^2\eta^2+k^2)^2}\frac{1}{k^3} < \frac{\varepsilon}{\eta^{3/2}}\frac{K}{i^{5/2}},$$

$$\sum_{k_o}|b_{ik}^{(2)}C_k| < \varepsilon^2\frac{K}{i^{5/2}},$$

and in general

$$\sum_k|b_{ik}^{(p)}C_k| < \varepsilon^p\frac{K}{i^{5/2}}, \quad p \text{ even},$$

$$< \frac{\varepsilon^p}{\eta^{3/2}}\frac{K}{i^{5/2}}, \quad p \text{ odd}.$$

Hence the series corresponding to (25) is absolutely convergent.

If η is so small that $\varrho(\pi\eta)$ is greater than 1, take out the factor $\varrho(\pi\eta)$ from each sum taken with respect to an odd second index. The result is that in case i is odd the factor $\varrho(\pi\eta)$ will appear in the first and second inequalities, the factor $[\varrho(\pi\eta)]^2$ in the third and fourth, etc. Consequently

$$\sum_k|b_{ik}C_k| + \sum_k|b_{ik}^{(2)}C_k| + \cdots < \frac{K\varepsilon\varrho(\pi\eta)(1+\varepsilon\eta^{3/2})}{i^{5/2}\eta^{3/2}}\{1+\varepsilon^2\varrho(\pi\eta)+\varepsilon^4[\varrho(\pi\eta)]^2+\cdots\}.$$

This series is convergent if $\varepsilon^2\varrho(\pi\eta)<1$, i. e., if $\varrho(\pi\eta)<2.198$. This is the case if $\eta>0.15$. This means that convergence is assured for most cases of practical interest.

When i is even the same methods apply. It is to be observed that the factor $i^{5/2}$ which appeared in the denominator in the foregoing inequalities is now replaced by $(i-1)^{5/2}$.

That (25) is a solution follows as in the case of the square. The proof of the uniqueness of the solution (25) requires that as in the case of the square

$$\lim_{p=\infty}\sum_k|b_{ik}^{(p)}| = 0.$$

This can readily be shown to be true if

$$\epsilon^2 \varrho(\pi\eta) < 1, \text{ i. e., if } \eta > 0.15.$$

It has been shown that if i is odd

$$|x_i| = |a_i| < \frac{M}{i^{5/2}},$$

if i is even

$$|x_i| = |b_{i-1}| < \frac{M}{(i-1)^{5/2}}, \quad \text{i. e.,} \quad |b_m| < \frac{M}{m^{5/2}}.$$

The present method of obtaining solutions of the system (27) is not established for η less than 0.15, i. e., for extremely narrow rectangles. As in most cases of practical interest η is greater than 0.15 it does not seem worth while to attempt to modify the proof to include the case of these small values of η.

Again it appears from the rapid decrease of the values of a_m and b_m with increasing m, that the system (16) and (17) may be solved by first taking n equal to 1 and omitting all terms on the right in which a and b occur with an index greater than 1. Thus approximate values of a_1 and b_1 are obtained. These values may then be entered in a second pair of equations obtained by taking $n = 3$ and omitting all terms after those in a_3 and b_3. The approximate values of a_3 and b_3 thus obtained may be used in the first pair of equations to secure more nearly correct values of a_1 and b_1. The process may be continued indefinitely. As it appears that in general only the terms in a_1 and b_1 need be retained in (7), it is necessary to calculate only these quantities with great care. Below is given a table of the values of a_1, b_1, a_3, b_3, apart from the common factor a^3, calculated for various values of η by the method just described.

η	a_1	b_1	a_3	b_3
0.2	0.052	0.001	0.0017	0.00001
0.35	0.089	0.006	0.0020	−0.00008
0.5	0.122	0.019	0.0015	−0.00025
0.667	0.148	0.047	0.0005	−0.00055
0.8	0.161	0.084	−0.0003	−0.00081
0.9	0.165	0.121	−0.0007	−0.00098
1.0	0.167	0.167	−0.0010	−0.00103

UNIVERSITY OF WISCONSIN,
MADISON, WIS.

ON IRREDUNDANT SETS OF POSTULATES*

BY

ALONZO CHURCH

1. **Definitions.** We shall say that the postulate A is weaker than the postulate B if B implies A and A does not imply B. This is a definition of the relative strength of two postulates in an absolute sense, as distinguished from their relative strength in the presence of other postulates.

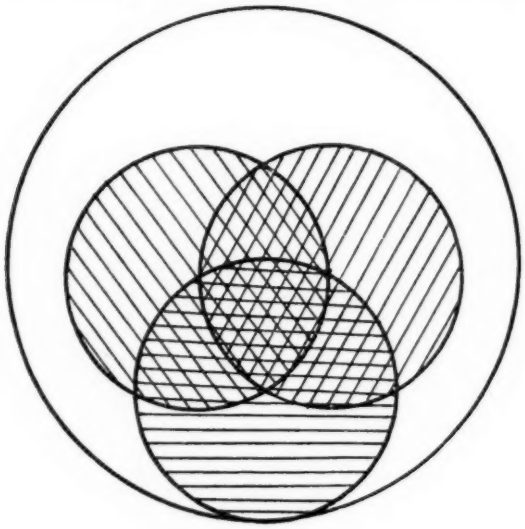


Fig. 1

The postulate A_1 can be weakened with respect to the set $A_1, A_2, A_3, \cdots, A_n$ if there exists a postulate, A_1', weaker than A_1, such that the logical product $A_1' A_2 A_3 \cdots A_n$ implies A_1. Under this definition we say that a postulate can be weakened with respect to a set of postulates, of which it is one, only when it can be replaced in the set by a weaker postulate in such a way that none of the implications of the set as a whole is lost.

A set of postulates is *irredundant* if the postulates are independent and no one of them can be weakened with respect to the set.

* Presented to the Society, December 30, 1924.

2. **The relation of irredundance to complete independence.*** If we represent the set of all possible mathematical systems by the set of all points in the interior of a circle, we may represent a postulate by a curve which divides the interior of the circle into two regions, which represent, one the set of systems which satisfy the postulate, and one the set of those which do not. We may distinguish between these two regions by shading the one within which the postulate is false and leaving the other unshaded. It is then clear that, if the postulate A is weaker than the postulate B the shaded area corresponding to A in the diagram will lie entirely within that corresponding to B.

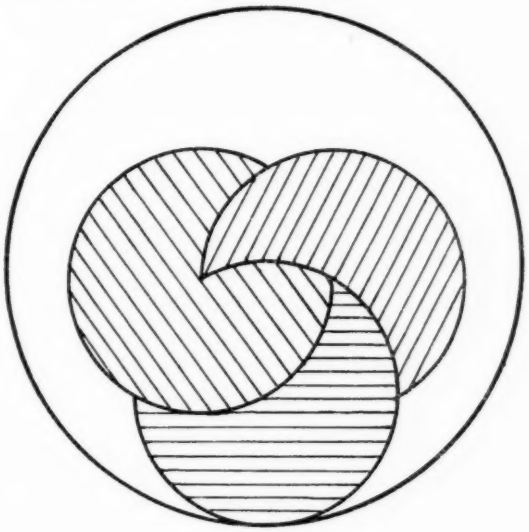


Fig. 2

In this way we may represent a set of three completely independent postulates as shown in Figure 1. The postulates are completely independent because, as is easily verified from the figure, systems exist which satisfy some of the postulates and fail to satisfy the others in any arbitrarily assigned manner. This corresponds to the fact that the three small circles which represent the postulates divide the interior of the large circle into 2^3, or 8, regions.

* The notion of complete independence is due to E. H. Moore. See *Introduction to a Form of General Analysis, New Haven Colloquium*, 1906, p. 82. For a discussion of the significance of this notion, see E. V. Huntington, *A new set of postulates for betweenness, with proof of complete independence*, these Transactions, vol. 26 (1924), p. 277.

If this set is to be made irredundant, the postulates must be weakened until no two of the shaded regions overlap; for so long as two of the shaded regions do overlap, it is possible to shrink one of them without reducing the total shaded area. The weakened set of postulates is represented in Figure 2. On the other hand, if a set of postulates which are not completely independent are to be made so, it may be necessary to strengthen one or more of them.* In particular, if the set is irredundant, all but one of the postulates, at least, must be strengthened.

A set of n completely independent postulates divides the set of all possible systems into 2^n parts, represented in a diagram by regions, the greatest number possible. The division effected by an irredundant set of n postulates, if the set is consistent, is a division into $n+1$ parts, the least number which is possible if the postulates are to be consistent and independent.

In general, an irredundant set of postulates is as far removed as possible from a completely independent set.†

3. **A criterion for irredundance.** In the diagram just described the relation of implication between postulates is represented by the relation of inclusion between regions. Thus, if A implies B, the shaded area corresponding to A contains that corresponding to B, and if the negative of A implies B, the unshaded area corresponding to A contains the shaded one corresponding to B. The observations made in the preceding section, therefore, suggest the following theorem:

A necessary and sufficient condition that the postulate A_1 cannot be weakened with respect to the set $A_1, A_2, A_3, \cdots, A_n$ is that the negative of A_1 imply each of the postulates $A_2, A_3, \cdots, A_n$.

The condition is necessary. For let B stand for the logical product $A_2 A_3 \cdots A_n$, and suppose that *not*-A_1 does not imply B. Then *not*-B does not imply A_1, and, therefore, A_1 *or* *not*-B does not imply A_1. The postulate A_1 *or* *not*-B is, therefore, weaker than A_1, and can be substituted for A_1 in the set $A_1, A_2, A_3, \cdots, A_n$, without loss of any of the implications of the set.

The condition is sufficient. For suppose that *not*-A_1 implies B, but that there exists a postulate A_1', weaker than A_1, such that the logical product

* For an interesting example of this, see J. S. Taylor, *Complete existential theory of Bernstein's set of four postulates for Boolean algebras*, Annals of Mathematics, ser. 2, vol. 19 (1917), p. 68, and the footnote on that page.

† The idea of a set of mutually prime postulates as defined by H. M. Sheffer is similar to that of an irredundant set as here defined. The two ideas are not, however, the same, because mutual primeness implies complete independence in the sense of E. H. Moore. See an abstract of a paper by H. M. Sheffer, *Mutually prime postulates*, Bulletin of the American Mathematical Society, vol. 22 (1916), p. 287. See also *The General Theory of Notational Relativity*, a privately circulated paper.

$A_1' B$ implies A_1. Then since *not*-A_1 implies B, it follows that *not*-B implies A_1; but $A_1' B$ implies A_1, and, by the law of excluded middle, B or *not*-B; therefore A_1' implies A_1, contrary to the hypothesis that A_1' was weaker than A_1.

The validity of this theorem depends on the validity of certain postulates of logic on which it is based, among which, as we have just seen, is the law of excluded middle. The conception with which we began, that a postulate divides the set of all possible mathematical systems into two parts, also depends, of course, on this law.

From the theorem which has just been proved follows as a corollary

A necessary and sufficient condition that a set of postulates be irredundant is that the postulates be independent and that the negatives of every two be contradictory.

Thus if the independent postulates A, B, C are to be irredundant, we must have *not*-A incompatible with *not*-B, *not*-A incompatible with *not*-C, and *not*-B incompatible with *not*-C.

There is a mechanical method by which any set of independent postulates can be made irredundant. If A, B, C, D are independent postulates, the set A, B', C', D' is an equivalent set which is irredundant, if B' is the proposition *if A then B*, and C' is the proposition *if A and B then C*, and D' the proposition *if A and B and C then D*. This method is open to the objection that it is likely to lead to postulates very complicated in form, and that, in general, the hypothesis of a postulate obtained in this way will be concerned with elements which are not mentioned in the conclusion. For example, if we applied this method to Peano's postulates for the system of positive integers*, we should obtain as our second postulate the following; "If 1 is a number, then every number has a unique successor."

The question arises to what extent it is possible to obtain irredundant sets of postulates which are not open to this objection. As a partial answer to this question, we shall give two examples of such sets of postulates.

4. **A simple example of an irredundant set.** The following set of postulates describes a system of a finite number of elements, which we call numbers, arranged in cyclic order. The undefined terms are *number* and *successor*. The postulates are the following:

1. There is a set of numbers, not null, such that every number in the set has a successor which is in the set.

2. In every proper part P, not null, of the set of all numbers there is a number none of whose successors is in P.

* Rivista di Matematica, vol. 1 (1891), pp. 87–102. See also below.

The negatives of these two postulates are

Not-1. There is no set of numbers, other than the null set, such that every number in the set has a successor which is in the set.

Not-2. In some proper part P, not null, of the set of all numbers every number has a successor which is in P.

Since these two statements are plainly contradictory, it will follow, as soon as we give independence examples, that the set of postulates 1 and 2 is irredundant.

An independence example for the first postulate is a system of a finite number of elements arranged in linear order. An independence example for the second postulate is the system of positive integers.

In the case of each example, when we have verified that it fails to satisfy the postulate for which it is given as an independence proof, we need not afterwards verify that it satisfies the other postulate, because we know that the negatives of the two postulates are contradictory. This corresponds to the fact that if a set of postulates satisfy the condition of the preceding section, that is, if the negatives of every two be contradictory, then a postulate of the set can fail to be independent only if its negative be self-contradictory in the system of logic on which the postulates are based.

From postulate 1 above follows the existence of at least one number. From postulate 2 it follows that the set whose existence is required by postulate 1 is the set of all numbers, and, therefore, that every number has a successor.

If a is any number, every number can be reached from a by proceeding from number to successor a finite number of times, for if A is the set of all numbers which can be so reached from a, it is true that every number in A has a successor in A, and therefore, by postulate 2, that A is the set of all numbers. If any number is successor of itself, it follows from postulate 2 that it is the only number, and in this case we have a cycle which contains just one element. In any other case we can find distinct numbers, a and b, such that b is a successor of a. Then there are one or more routes by which a can be reached from b in a finite number of steps from number to successor. And it follows from postulate 2 that any arbitrarily chosen one of these routes must contain every number. The set of all numbers is, therefore, finite, and can be arranged in a cycle in such a way that every number is a successor of the number which next precedes it in the cycle.

Finally, there is no number a which has two successors, b and c. For, if there were, we could find a finite route from b to c, and this route would not contain a, because every finite route from b to a contains all numbers. Similarly there would be a finite route from c to b which did

not contain a, and the set P of all numbers which were in either of these routes would have the property that every number in it had a successor in it. But P would not contain a, and would, therefore, violate postulate 2.

We have, therefore, shown that postulates 1 and 2 uniquely determine a system of a finite number of elements arranged in cyclic order.

5. **An irredundant set of postulates for the system of positive and negative integers.*** The following postulates, which deal with a set of undefined elements, *numbers*, and an undefined relation, *successor* among them, describe the system of positive and negative integers:

1. There is a set of numbers, S, not null, such that every number in S has a successor which is in S.

2. If the set of all numbers is such a set S, it is not the only such set.

3. If such sets S exist, one of them has the property that every number in the set is successor of some number in the set.

4. In every proper part P, not null, of the set of all numbers, either some number has no successor which is in P, or some number is successor of no number in P.

The negatives of these postulates are:

Not-1. There is no set of numbers S, other than the null set, such that every number in S has a successor which is in S.

Not-2. The set of all numbers is an S and the only S.

Not-3. Sets S exist, but no one of them has the property that every number in the set is successor of some number in the set.

Not-4. There is a proper part P, not null, of the set of all numbers, such that every number in P has a successor in P and is successor of some number in P.

Of these four statements every two are contradictory. This is evident at once, except in the case of *not*-2 and *not*-3. In order to see that *not*-2 and *not*-3 are contradictory, we observe that if the set of all numbers, N, were an S and the only S, then every number in N would be successor of some number in N, because if the number a in N were successor of no number in N we could, by omitting a from N, obtain an S not the set of all numbers.

Independence examples are as follows. For postulate 1, a system of a finite number of elements arranged in linear order; for postulate 2, a finite

* For a set of independent postulates for the system of positive and negative integers, see A. Padoa, *Un nouveau système irréductible de postulats pour l'algèbre*, Deuxième Congrès International des Mathématiciens, Paris, 1900, pp. 249–256, and *Numeri interi relativi*, Rivista di Matematica, vol. 7 (1901), pp. 73–84. This set is, however, not categorical, because it is satisfied by the set of integers reduced modulo m, where m is any odd integer.

cycle; for postulate 3, the system of positive integers; for postulate 4, a system of two sets of a finite number of elements each, no element common to the two sets, and each set arranged separately in cyclic order. As pointed out in the preceding section, it is sufficient to see, in the case of the first example, that it satisfies *not*-1, because every system which satisfies *not*-1 must also satisfy postulates 2, 3, and 4; and similarly in the case of each of the other examples.

The set of postulates 1, 2, 3, 4 is, therefore, irredundant.

6. **A proof that the foregoing set of postulates is categorical.** We shall prove that the postulates given above uniquely determine the system of positive and negative integers. In order to do this, we shall show that the set of all numbers, as defined by these postulates, can be divided into two parts which have in common one and only one number, a, of which one satisfies Peano's postulates for the system of positive integers* when "1" is replaced in them by "a", and the other satisfies Peano's postulates when "1" is replaced in them by "a" and "successor" by "predecessor".

Peano's postulates are as follows:

1. 1 is a number.

2. Every number has a unique successor.

3. If a and b are numbers, and if their successors are equal, then a and b are equal.

4. 1 is not the successor of any number.

5. If S is a class of numbers which contains 1, and the class of all successors of numbers in S is contained in S, then every number is contained in S.

If not imperative, it is at least desirable that the proposed proof should be made without the use of the notion of finite cardinal number, although this necessitates an argument somewhat longer than might otherwise be given.

THEOREM 1. *There exists a number.* (Postulate 1 of § 5.)

THEOREM 2. *Every number has a successor.*

For it follows from postulate 4 that the set whose existence is required by postulate 3 is the set of all numbers.

DEFINITION. If the number a is successor of the number b, then b is *predecessor* of a.

THEOREM 3. *Every number has a predecessor.* (Postulates 1, 3, 4.)

THEOREM 4. *No number is successor of itself.*

For if a were successor of itself, then a would be the only number, by postulate 4. But in that case postulate 2 would be violated.

* Loc. cit. See also Formulaire de Mathématiques, vol. 3 (1901), pp. 39–44.

THEOREM 5. *If the number b is successor of the number a, a is not successor of b.*

For if a were successor of b, a and b would be the only numbers, by postulate 4. But in that case postulate 2 would be violated.

DEFINITION. The set of numbers Q is an *interval* if it contains numbers, a and b, which have, respectively, no predecessor and no successor in Q, and every number other than a has a unique predecessor in Q, and every number other than b has a unique successor in Q, and finally, there is no part of Q, not null, in which every number has a successor. The numbers a and b are the *end numbers* of Q, and Q is *an interval* ab.

THEOREM 6. *The set of numbers obtained by omitting b from an interval ab, if $b \neq a$, is an interval ac, where c is the predecessor of b in ab. The set of numbers obtained by omitting a from an interval ab, if $b \neq a$, is an interval db, where d is the successor of a in ab.*

THEOREM 7. *If intervals ab and bc have the property that no number in either, other than b, has a successor in the other, other than b, then the sum, P, of ab and bc is an interval ac.*

Suppose that in some part T of P, not null, every number had a successor. Then the set R of numbers common to T and bc would have the property that every number in it had a successor in it, for otherwise some number in ab would be successor of some number in bc. But R cannot be null, because, if it were, I would be a part of ab. Therefore there is no part of P in which every number has a successor. Therefore, since it is easily seen that P has the other properties required by the definition, it follows that P is an interval ac.

COROLLARY 1. *If c is a successor of b, and, in the interval ab, c has no successor and no predecessor other than b, then the set of numbers obtained by adding c to ab is an interval ac.*

COROLLARY 2. *If c is a predecessor of a, and, in an interval ab, c has no predecessor and no successor other than a, then the set of numbers obtained by adding c to ab is an interval cb.*

THEOREM 8. *If c is a number of an interval ab, this interval can be divided into intervals, ac and cb, whose sum is ab, and which have only c in common.*

Let P be the set of all numbers, c, in ab of which this is not true. If the theorem is true of any number, d, in ab, it is true also of the predecessor of d in ab (Theorem 6 and Theorem 7, Corollary 1). Therefore, if any number c is contained in P, the successor of c in ab is also contained in P. This is true without exception, since b is not contained in P, because ab can be divided into ab and bb, where bb consists of b alone. Therefore, by the definition of interval, the set P is null.

THEOREM 9. *This division is unique.*

For suppose there were another such division of ab, say a division into $(ac)'$ and $(cb)'$. Let P be the set of all numbers in ac and not in $(ac)'$. Since every number in ac, other than c, has a successor in ac, and every number in $(ac)'$, other than a, has a predecessor in $(ac)'$, it follows that if P contains d it contains also the successor of d in ab. Therefore P is null. And, therefore, every number in ac is contained also in $(ac)'$. By the same reasoning, every number in $(ac)'$ is contained also in ac. Therefore ac and $(ac)'$ are identical. Therefore cb and $(cb)'$ are identical.

DEFINITION. The number d, in ab, different from c, in ab, *precedes* or *follows* c *in* ab according as d is in ac or cb.

THEOREM 10. *If d precedes c in ab, then c follows d in ab; and if c follows d in ab, then d precedes c in ab.*

For by Theorems 7, 8, 9, we have

$$ab = ac+cb = ad+dc+cb = ad+db,$$

where db contains c.

THEOREM 11. *If c_1 precedes c_2 in ab and c_2 precedes c_3 in ab, then c_1 precedes c_3 in ab.*

For we have

$$ab = ac_1+c_1b = ac_1+c_1c_2+c_2b = ac_1+c_1c_2+c_2c_3+c_3b = ac_1+c_1c_3+c_3b,$$

and since $c_1c_3+c_3b = c_1b$, by Theorem 7, it follows that c_1b contains c_3.

THEOREM 12. *In any subset P, not null, of an interval ab there is a first number and a last number.*

Let p stand for any number in P. Then in the set R of all numbers which occur in any interval ap, there is a number, c, which has no successor in the set. This number c is a member of P. For, if it were not, it would be a number, not an end number, in some interval ap, and its successor in ap would be a member of R. And there is no number p, in P, which follows c in ab. For, if there were, c would be contained in ap, and its successor in ap would be a member of R. Finally, if d is any number in P, different from c, d must precede c in ab, for c does not precede d, and it follows at once from the definition that one of the numbers c and d precedes the other in ab. Therefore c is the last number in P.

In the set T of all numbers which occur in every interval ap is a number, e, which has no successor in the set. This number e is a member of P. For, if it were not, it would be a number, not an end number, in every interval ap, and its successor in ab would be in every ap, and would, therefore, be a member of T. And there is no number p, in P, which precedes e in ab. For, if there were, e would not be contained in ap, and

would, therefore, not be a member of T. And if d is any number in P, different from e, d must follow e in ab. Therefore e is the first number in P.

THEOREM 13. *If a and b are distinct numbers, there do not exist both an interval ab and an interval ba.*

Suppose both do exist. Then, by postulate 4, $ab+ba$ is the set of all numbers.

Suppose that, apart from a and b, ab and ba have no number in common, and no number in either is successor of any number in the other. Then, by postulate 2, there is some proper part, T, of $ab+ba$ in which every number has a successor. There is a number, c, in $ab+ba$ which is not contained in T. It is possible to choose c distinct from a and b. For, if a is not in T, the predecessor of a in ba is also not in T, and if b is not in T, the predecessor of b in ab is also not in T; and since, by Theorem 5, either ab or ba must contain numbers other than a and b, we can always find a number distinct from a and b and not in T, as required. Let c be so chosen, and suppose that c is in ab. Let d and e be the predecessor and the successor, respectively, of c in ab. Then $eb+ba+ad$ is an interval ed by Theorem 7. But this interval contains a part, T, in which every number has a successor, contrary to the definition of interval.

Suppose that, apart from a and b, ab and ba have no number in common, and that some number in one of them, different from a and b, say c in ab, has a successor in the other different from a and b, say d in ba. Then $ac+da$ is a proper part of the set of all numbers in which every number has a successor and a predecessor, contrary to postulate 4.

Suppose that ab and ba have a number, c, in common, other than a and b. Then $ac+ca$ is a proper part of the set of all numbers in which every number has a successor and a predecessor, contrary to postulate 4.

Therefore, in every case, the supposition that both ab and ba exist leads to a contradiction.

THEOREM 14. *If c is a successor of b, the existence of an interval ab implies the existence of an interval ac.*

The number c has no successor, d, in ab. For, if it had, either d would be the predecessor of b in ab, in which case b, c, and d would be the only numbers, by postulate 4, and postulate 2 would be violated, or, if d were not the predecessor of b, the numbers c and d would constitute an interval cd and c together with db would constitute an interval dc, contrary to the preceding theorem. Now there are numbers in ab of which c is successor, for b is such a number. Let d be the first such number (Theorem 12). Then ad and c together constitute an interval ac, by Theorem 7, Corollary 1.

THEOREM 15. *If c is a predecessor of a, the existence of an interval ab implies the existence of an interval cb.*

This follows by reasoning similar to the preceding, using Theorem 7, Corollary 2, in place of Theorem 7, Corollary 1.

DEFINITION. If a is any number, the *ray* A is the set of all numbers b such that an interval ab exists, and the *ray* A' is the set of all numbers c such that an interval ca exists.

From the theorems which precede, we can infer at once that the ray A contains a, and contains no predecessor of a, that every number in A has a successor in A, and that every number but a in A has a predecessor in A. Similarly, the ray A' contains a, and contains no successor of a, and every number in A' has a predecessor in A', and every number but a in A' has a successor in A'.

THEOREM 16. *The rays A and A' have no number other than a in common.*

For, if they had b, different from a, in common, intervals ab and ba would both exist, contrary to Theorem 13.

THEOREM 17. *If the numbers b and c are successors of the number a, then b and c are the same number.*

For suppose that b and c were distinct. Then, if the ray B did not contain c, $A' + B$ would be a proper part of the set of all numbers of which postulate 4 would not be true. Therefore B contains c, and, for the same reason, C contains b. But neither B nor C contains a. Therefore $B + C$ is a proper part of the set of all numbers in which every number has a successor and a predecessor, contrary to postulate 4.

THEOREM 18. *If the number a is successor of the numbers b and c, then b and c are the same number.*

THEOREM 19. *$A + A'$ is the set of all numbers.* (Postulate 4.)

THEOREM 20. *If S is a part of the ray A, such that a is a member of S, and the class of all successors of numbers in S is contained in S, then S contains every number of A.*

For otherwise $S + A'$ would be a proper part of the set of all numbers in which every number had a successor and a predecessor.

THEOREM 21. *If S is a part of the ray A', such that a is a member of S, and the class of all predecessors of numbers in S is contained in S, then S contains every number of A'.*

For otherwise $S + A$ would violate postulate 4.

We have now shown that the set of all numbers can be divided into two parts, A and A', which have a only in common and which satisfy Peano's postulates in the manner described above. Therefore the set of postulates 1, 2, 3, 4 with which we began is categorical.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

THREE-DIMENSIONAL MANIFOLDS OF STATES OF MOTION*

BY

HAROLD HOTELLING

1. Professor Birkhoff has shown† that every dynamical problem with two degrees of freedom can be represented by the motion of a particle under a field of force on a characteristic surface which is either fixed (reversible case) or rotating at uniform velocity and carrying its field of force with it (irreversible case). In either type of problem the magnitude of the velocity is a function only of the position of the particle on the surface, so that the state of motion at any instant is given by three parameters, two giving the position and one the direction of motion. We shall for the most part consider the last named parameter as varying from 0 to 2π, regarding opposite directions at a point as distinct; this is essential in the irreversible though not in the reversible case.

The three parameters may be regarded as determining a point in a three-dimensional manifold. The manifolds thus obtained are considered in this paper from the point of view of analysis situs.

It should be noted that these manifolds are not in general, as might be expected, the "product manifolds" whose points are in 1 – 1 continuous correspondence with pairs of points, one point of each pair lying on the characteristic surface and one on a circle. For example, if the characteristic surface is one-sided the product manifold is one-sided, but the manifold of states of motion is orientable. Only for a surface of the connectivity of the anchor ring are the two manifolds homeomorphic. The underlying reason is the impossibility of a non-singular coördinate system or continuous distribution of vectors on any other surface.

Each orbit is represented by a curve in the manifold. These curves are called "stream lines" by Birkhoff (loc. cit.) because of a hydrodynamical analogy. The stream lines constitute a non-singular congruence, one through each point of the manifold. To a periodic orbit corresponds a closed stream line. This relation and the importance of periodic orbits add interest to the problem of finding the properties of the manifolds of states of motion. Some special manifolds of this class are described, from a point of view slightly different from that here adopted, by Birkhoff in the paper above

* Presented to the Society, San Francisco Section, October 25, 1924.

† *Dynamical systems with two degrees of freedom*, these Transactions, vol. 18 (1917), pp. 212 et seq.

cited and in a paper on *The restricted problem of three bodies* in the Rendiconti del Circolo Matematico di Palermo, vol. 39 (1915), pp. 265–334.

We limit our attention to cases in which the functions appearing in the differential equations of motion are finite throughout the region considered.

2. We shall make use of the *Heegaard diagram*. This method of representing a 3-dimensional manifold supposes it divided by a surface of genus p into *canonical regions*,* each of which is characterized by the fact that it can be represented by the interior of a sphere with p non-linking and knotless handles in euclidean 3-space. A canonical region can be made simply connected by the removal of p 2-cells, one severing each handle. We select a particular set of 2-cells having this property and call them *canonical cuts* and their boundaries *canonical curves*. The regions are sometimes designated as the "red" and "green" regions and their respective canonical curves as the "red" and "green" curves. The canonical curves are evidently not deformable to a point on the common surface of the two regions. If the red curves are traced on the surface of the representation in euclidean space of the green region we have a diagram which completely characterizes the manifold.

The *group* of the manifold may be obtained from the group of the surface as follows. Let the generators of the latter group be a set of p green curves $g_1, g_2, \cdots, g_p$ and p other curves $a_1, a_2, \cdots, a_p$ such that a_i and g_i $(i = 1, 2, \cdots, p)$ intersect once. Relations (which may be redundant) among the a's are found by expressing each of the red curves $r_1, r_2, \cdots, r_p$ in terms of the g's and a's and then putting each g and each r equal to the identity. The group of the 3-dimensional manifold is generated by $a_1, \cdots, a_p$, subject to the relations just found.

3. **The sphere.** To obtain the manifold of states of motion on a surface of genus 0, represent the surface by a sphere, divide the latter by a great circle, and map each hemisphere conformally on the interior of a circle. The part of the manifold corresponding to each of these regions will be homeomorphic with the interior of a tore, since a vector at the point (x, y) of the circle or its interior making an angle φ $(0 \leqq \varphi < 2\pi)$ with the positive x-axis may be represented by the point (x, y, φ) in space; the cylinder thus obtained is to be deformed so that the face $\varphi = 0$ coincides with the face $\varphi = 2\pi$. We choose $\varphi = 0$ as canonical cut in each

* This and the other terms introduced in this section have been used in a work as yet unpublished by Professor J. W. Alexander, to whom is due also this method of exposition of the Heegaard diagram. A proof that a division of the manifold into two canonical regions is always possible will be found in the 1916 *Cambridge Colloquium Lectures*, by O. Veblen, p. 148 (New York, 1922).

case. Let us denote points on one circle by the coördinate θ measuring the angle made by the radius at the point with the positive x-axis, and let us denote the corresponding points on the other circle by θ'. Let the mapping be such that θ' is the angle which the radius of the second circle makes with the positive x-axis. Since a vector at the common boundary of the two parts of the surface makes the same angle ψ with the inward-drawn normal in one case as with the outward-drawn normal in the other, the equations expressing the transformation of one tore into the other are (Fig. 1)

$$\varphi' = 2\theta - \varphi + \pi, \tag{1}$$
$$\theta' = \theta.$$

Along the canonical curve for one region – the red region, say—φ is zero and θ varies through 2π. Hence, by (1), φ' varies through 4π and

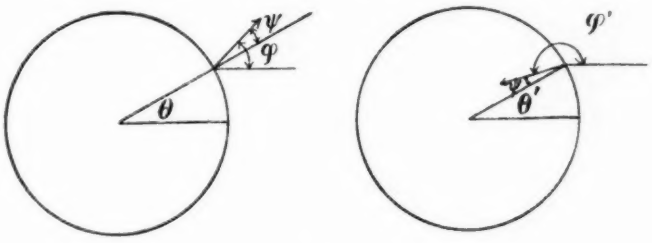


Fig. 1

θ' through 2π. This red curve cuts the green curve $\varphi' = 0$ twice; the Heegaard diagram may be taken as a tore and its interior, representing the green region, the red curve being traced on the surface. The red curve will go, so to speak, twice *around* the hole in the doughnut and once *through* the hole. Denoting a generating circle of this tore by g and a curve which cuts it once by a, the red curve is denoted by $r = a^2 g$. Hence the fundamental group of the manifold, obtained by putting $r = 1$, $g = 1$, is cyclic and of order 2.

This manifold is homeomorphic with the projective 3-space. Indeed the latter can be divided by a one-sheeted hyperboloid of revolution into two regions, each of which is homeomorphic with the interior of a tore. A plane through the axis of revolution cuts the surface in a hyperbola which bounds a 2-cell in the region not containing the center. This hyperbola has the same relation to the region containing the center that the curve on the Heegaard diagram described above bears to the region in the diagram. Having the same Heegaard diagram the two manifolds are homeomorphic.*

* O. Veblen, *Cambridge Colloquium Lectures*, p. 149.

The manifold is also homeomorphic with a two-sheeted Riemann space whose branch system consists of a pair of linking circles AGH and CDE (Fig. 2). This will appear if a pair of cuts separating the sheets is made along a strip generated in a non-singular manner by a straight segment joining a point of one circle to a point of the other, the segment moving in such a way that each end sweeps out one of the circles. The two strips form a tore. Each sheet is then a canonical region of genus 1, for it is bounded by this tore and is made simply connected by the removal of a 2-cell whose boundary is $ABCDEFGHA$ in one case, $AB'CDEF'GHA$ in the other; we suppose here that ABC and EFG lie in one sheet and $AB'C$ and $EF'G$ in the other. These curves may be deformed on the tore until they intersect in only two points, and this is the minimum number of intersections. The Heegaard diagram obtained in this way is the same as that found above for the manifold of states of motion.

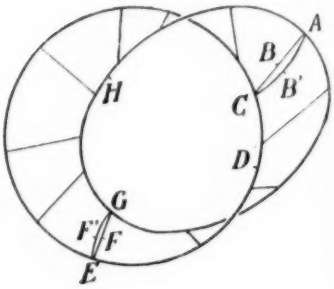


Fig. 2

4. If opposite directions are considered identical, φ will vary only from 0 to π. The cylinder mentioned in § 3 will now have altitude π. Equations (1) still hold, showing that as θ varies from 0 to 2π, φ being constant, φ' varies through 4π. Hence the Heegaard diagram will be a tore on which the characteric curve is a^4g. The group is cyclic and of order 4.

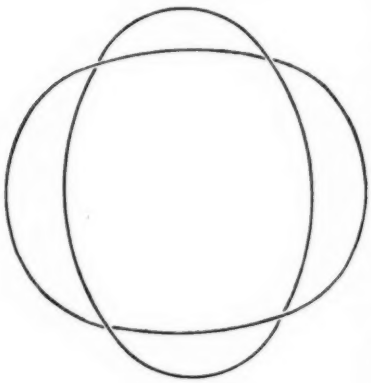


Fig. 3

The manifold is homeomorphic with a two-sheeted Riemann space whose branch system consists of a pair of knotless curves linking each other doubly in the manner indicated in Fig. 3. This is shown by an argument similar to that of the last paragraph of § 3.

5. **The projective plane.** A one-sided surface of connectivity 2 is representable by a circle and its interior, diametrically opposite points of the circle being considered identical. A vector on the boundary making an angle ψ with the outward-drawn normal is therefore to be identified with a vector at the diametrically opposite point which makes an angle $-\psi$ with the inward-drawn normal. Using the polar coördinate θ (pole at the

center) to represent points on the boundary, it is evident from principles of elementary geometry that a vector at θ making an angle φ with the direction of the polar axis is identical with a vector at $\theta + \pi$ making an angle $2\theta - \varphi$ with this direction.

The manifold of states of motion can therefore be represented by a cylinder and its interior, the ends of the cylinder being considered identical and the points of the curved surface being matched in such a way that (θ, φ) is identical with $(\theta + \pi, 2\theta - \varphi)$. Thus the points $(0, 0)$, $(\pi, 0)$, $(0, 2\pi)$ and $(\pi, 2\pi)$ all represent the same point A of the manifold (Fig. 4). Let α be a curve on the cylinder for which $\varphi = 0$ and $0 \leq \theta \leq \pi$. Then α represents a curve in the manifold which is also represented on the cylinder by $\varphi = 2\pi$, $0 \leq \theta \leq \pi$ and by $\varphi \equiv 2\theta \pmod{2\pi}$, $\pi \leq \theta \leq 2\pi$. Similarly, let β be a curve on the cylinder for which $\varphi = 0$ and $\pi \leq \theta \leq 2\pi$. It represents a curve in the manifold which is also represented by $\varphi = 2\pi$, $\pi \leq \theta \leq 2\pi$, and by $\varphi = 2\theta$, $0 \leq \theta \leq 2\pi$.

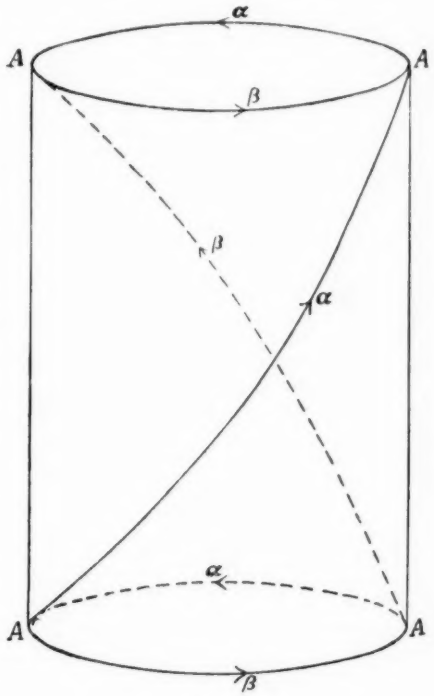


Fig. 4

To obtain a Heegaard diagram, surround* each of the curves in the manifold represented by α and β by a fine tube terminating in a ball surrounding A. On this ball with two handles are to be traced a pair of curves which bound 2-cells in the remainder of the manifold. For these 2-cells we may take the parts exterior to our fine tube of (1) an end of the cylinder and (2) a surface within the cylinder bounded by the semicircles $0 \leq \theta \leq \pi$ on the bases and by the curves $\varphi = 2\theta$. Thus we find for the group of the manifold two generators α and β connected by the two relations $\alpha\beta = 1$ and $\alpha^3\beta^{-1} = 1$, from which it follows that the group is cyclic and of order 4.

* Veblen, loc. cit., p. 148, § 41.

The manifold is also homeomorphic with a two-sheeted Riemann space whose branch system consists of a pair of linking circles AGH and CDE (Fig. 2). This will appear if a pair of cuts separating the sheets is made along a strip generated in a non-singular manner by a straight segment joining a point of one circle to a point of the other, the segment moving in such a way that each end sweeps out one of the circles. The two strips form a tore. Each sheet is then a canonical region of genus 1, for it is bounded by this tore and is made simply connected by the removal of a 2-cell whose boundary is $ABCDEFGHA$ in one case, $AB'CDEF'GHA$ in the other; we suppose here that ABC and EFG lie in one sheet and $AB'C$ and $EF'G$ in the other. These curves may be deformed on the tore until they intersect in only two points, and this is the minimum number of intersections. The Heegaard diagram obtained in this way is the same as that found above for the manifold of states of motion.

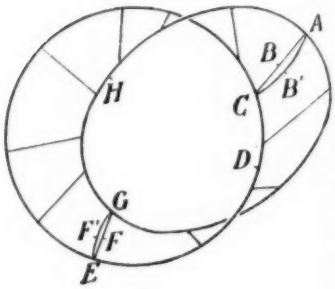


Fig. 2

4. If opposite directions are considered identical, φ will vary only from 0 to π. The cylinder mentioned in § 3 will now have altitude π. Equations (1) still hold, showing that as θ varies from 0 to 2π, φ being constant, φ' varies through 4π. Hence the Heegaard diagram will be a tore on which the characteric curve is a^4g. The group is cyclic and of order 4.

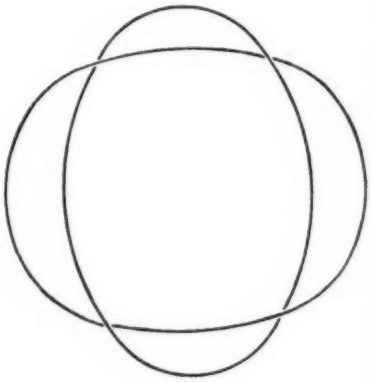


Fig. 3

The manifold is homeomorphic with a two-sheeted Riemann space whose branch system consists of a pair of knotless curves linking each other doubly in the manner indicated in Fig. 3. This is shown by an argument similar to that of the last paragraph of § 3.

5. **The projective plane.** A one-sided surface of connectivity 2 is representable by a circle and its interior, diametrically opposite points of the circle being considered identical. A vector on the boundary making an angle ψ with the outward-drawn normal is therefore to be identified with a vector at the diametrically opposite point which makes an angle $-\psi$ with the inward-drawn normal. Using the polar coördinate θ (pole at the

center) to represent points on the boundary, it is evident from principles of elementary geometry that a vector at θ making an angle φ with the direction of the polar axis is identical with a vector at $\theta + \pi$ making an angle $2\theta - \varphi$ with this direction.

The manifold of states of motion can therefore be represented by a cylinder and its interior, the ends of the cylinder being considered identical and the points of the curved surface being matched in such a way that (θ, φ) is identical with $(\theta+\pi, 2\theta-\varphi)$. Thus the points $(0, 0)$, $(\pi, 0)$, $(0, 2\pi)$ and $(\pi, 2\pi)$ all represent the same point A of the manifold (Fig. 4). Let α be a curve on the cylinder for which $\varphi = 0$ and $0 \leqq \theta \leqq \pi$. Then α represents a curve in the manifold which is also represented on the cylinder by $\varphi = 2\pi$, $0 \leqq \theta \leqq \pi$ and by $\varphi \equiv 2\theta \pmod{2\pi}$, $\pi \leqq \theta \leqq 2\pi$. Similarly, let β be a curve on the cylinder for which $\varphi = 0$ and $\pi \leqq \theta \leqq 2\pi$. It represents a curve in the manifold which is also represented by $\varphi = 2\pi$, $\pi \leqq \theta \leqq 2\pi$, and by $\varphi = 2\theta$, $0 \leqq \theta \leqq 2\pi$.

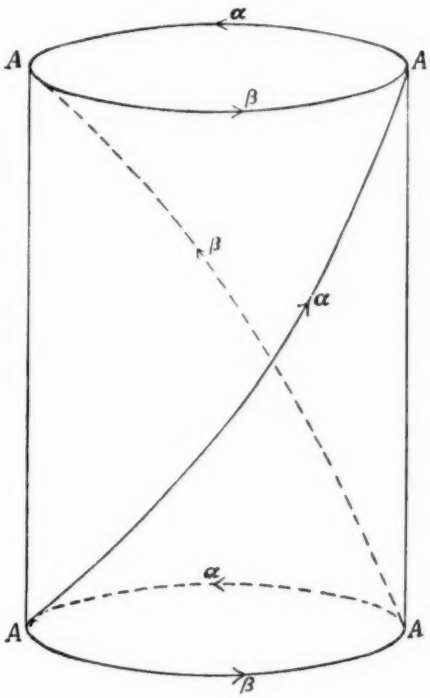


Fig. 4

To obtain a Heegaard diagram, surround* each of the curves in the manifold represented by α and β by a fine tube terminating in a ball surrounding A. On this ball with two handles are to be traced a pair of curves which bound 2-cells in the remainder of the manifold. For these 2-cells we may take the parts exterior to our fine tube of (1) an end of the cylinder and (2) a surface within the cylinder bounded by the semicircles $0 \leqq \theta \leqq \pi$ on the bases and by the curves $\varphi = 2\theta$. Thus we find for the group of the manifold two generators α and β connected by the two relations $\alpha\beta = 1$ and $\alpha^3\beta^{-1} = 1$, from which it follows that the group is cyclic and of order 4.

* Veblen, loc. cit., p. 148, § 41.

6. To show that this manifold is homeomorphic with that of § 4 we use a lemma due essentially to Heegaard:

If on a Heegaard diagram of genus 2 *a red curve* r *cuts a green curve* g *exactly once, the manifold may also be represented by a Heegaard diagram of genus* 1.

We may prove this lemma by representing the green region by a ball with two handles in such a way that g bounds a 2-cell severing one of the handles and any other curve g' which bounds a 2-cell in the interior but not on the surface and which does not intersect g and is not deformable into g bears the same relation to the other handle.

For the case in which r fails to meet g', so that it merely traverses the handle severed by the 2-cell bounded by g, the lemma may be proved simply by transferring from the red to the green region a 3-cell obtained by slightly thickening the 2-cell bounded by r.

A reduction to this case can always be effected on account of the freedom possible in choosing the second green curve g'. For we can always find a curve $g' = rgr^{-1}g^{-1}$ which is not deformable into g and which fails to bound on the surface but does bound a 2-cell in the green region. Such a 2-cell can be constructed by combining a strip bounded by a portion of r, counted twice, with a pair of 2-cells which are bounded by curves into which g is deformable on the surface. Since g' need not meet r this brings us to the former case.

7. Since one of the handles of the Heegaard diagram described in § 5 is traversed just once by either of the mentioned curves which bound in the other part of the manifold, it follows from the lemma of § 6 that the manifold has a Heegaard diagram consisting of a tore on which is traced one characteristic curve. Since the group is cyclic and of order 4 this curve must be expressible in the form $a^4 g^{2m+1}$. All the diagrams thus obtained are homeomorphic with a tore on which the characteristic curve is $a^4 g$.

The manifold of states of motion on a projective plane, opposite directions being considered distinct, is thus homeomorphic with the manifold for a sphere on which opposite directions are considered identical.

From this and from the theorem that on every smooth surface of genus 0 there is a closed geodesic it may be inferred that on every smooth surface of connectivity 2 there is a closed geodesic. This is otherwise evident because a two-sheeted covering surface of genus 0 may be placed upon a closed surface of connectivity 2.

8. If opposite directions on the projective plane are considered identical the manifold is represented by a cylinder whose bases correspond to each other and whose convex surface is divided by curves which correspond to

parts of the edges into four 2-cells which correspond to each other by pairs. By a method similar to that of § 5 it is easily found that the group of this manifold has two generators α and β connected by the two relations

$$\alpha\beta\alpha\beta^{-1} = 1 \qquad \text{and} \qquad \beta\alpha\beta\alpha^{-1} = 1.$$

It follows that the connectivity is 3 and that there are two coefficients of torsion, each equal to 2.

9. **Surfaces of the connectivity of the anchor ring.** Both the anchor ring and the one-sided surface of the same connectivity are loci of points on a circle which is deformed into itself through a non-singular set of intermediate positions. If the generating circle returns with sense preserved the surface is orientable; if with sense reversed it is non-orientable. In either case we set up a system of coördinates u and v such that u is constant along a generating circle, v is constant along a curve which cuts each generating circle once, $0 \leq u \leq 2\pi$ and $-1 \leq v \leq 1$. The point $(u, 1)$ coincides with $(u, -1)$. Also $(0, v)$ coincides with $(2\pi, v)$ if the surface is orientable, otherwise with $(2\pi, -v)$.

A direction at a point on the surface may be uniquely specified by the angle $\varphi\,(-\pi \leq \varphi \leq \pi)$ which it makes with the curve $v =$ constant through the point, angles being measured positively from the direction of increasing u to that of increasing v. φ varies continuously with a moving vector except, in the one-sided case, at $u = 2\pi$, where φ and v change sign abruptly.

The portion of the manifold of states of motion corresponding to a generating circle (or any other closed curve) is a tore, since its points can be specified by the two cyclic coördinates v and φ. The manifold can thus be represented by a continuous one-parameter family of coaxial tores $u =$ constant, each within all the preceding.

In the orientable case the innermost and outermost tores of the family are matched together in the manner determined simply by swelling the inner one: $(-1, v, \varphi)$ is identical with $(1, v, \varphi)$. But if the characteristic surface is one-sided the tores are to be matched in such a way that $(-1, v, \varphi)$ is identical with $(1, -v, -\varphi)$.

The manifold for the one-sided surface is homeomorphic with a two-sheeted Riemann space whose branch system consists of two pairs of circles, each pair forming the boundary of a plane ring and each curve linking once each curve of the other pair (Fig. 5). (These plane rings we take as the surfaces on which the two sheets join.) For if we construct in each sheet a pair of non-intersecting tores, one surrounding each of the plane rings, we divide the space into four parts, each of which is homeomorphic with the region between two coaxial tores of which one is within the other. Since each

part is the locus of points on a one-parameter family of non-intersecting tores, and since the four parts are arranged cyclically, the whole space is such a locus. The first and last tores of the family coincide with change of sign of both coördinates of a point, since one coördinate changes sign each time the moving tore passes from one sheet to the other and then covers a former position, and neither coördinate changes sign twice.

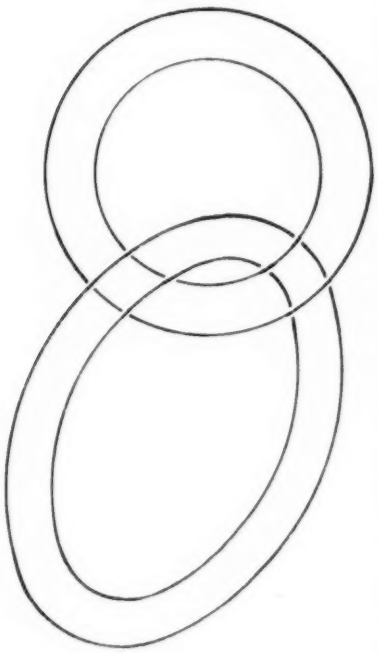


Fig. 5

In the same way the manifold of states of motion on a tore is seen to be homeomorphic with a four-sheeted space with the same branch curves if Sheet 1 permutes with Sheet 2 and Sheet 3 with Sheet 4 about each curve of one pair, while Sheet 1 permutes with Sheet 4 and Sheet 2 with Sheet 3 about each curve of the other pair.

10. **Closed surfaces in general.** We now pass to the general case of a closed surface of connectivity greater than that of the anchor ring. The letter k is used in this paper, whether referring to manifolds of two or of three dimensions, to mean the number of independent non-bounding closed curves. Thus for an orientable surface of genus p, $k = 2p$. The (linear) connectivity is $k+1$.

The surface may be represented* by a plane polygon, with its interior, having $2k$ directed edges which correspond by pairs, all the vertices representing the same point of the surface. In passing around the polygon the edges occur in the order

$$a_1\ b_1\ a_1^{-1}\ b_1^{-1}\ a_2\ \cdots\ a_p\ b_p\ a_p^{-1}\ b_p^{-1} \qquad (p = k/2)$$

if the surface is orientable; in the order

$$a_1\ b_1\ a_1^{-1}\ b_1^{-1}\ a_2\ \cdots\ a_p\ b_p\ a_p^{-1}\ b_p^{-1}\ d_1\ d_1 \qquad (p = (k-1)/2)$$

* Veblen, loc. cit., pp. 70-72, 137.

if the surface is one-sided and k is odd; or in the order

$$a_1\ b_1\ a_1^{-1}\ b_1^{-1}\ a_2\ \cdots\ a_p\ b_p\ a_p^{-1}\ b_p^{-1}\ d_1\ d_1\ d_2\ d_2 \quad (p=(k-2)/2)$$

if the surface is one-sided and k is even.

First suppose $k \geq 3$. We use a curvilinear polygon whose edges are congruent circular arcs making interior angles π/k at the vertices. Then since the sum of the interior angles is 2π the surface can be mapped conformally on the interior and boundary of the polygon. The angle subtended by each edge at its center is $m = \pi(1-2/k)$.

The possible states of motion at points within the polygon are represented by the points within a fluted prism of height 2π (regarding opposite directions as distinct) erected upon the polygon as base, the height of a point being equal to the angle φ which the corresponding vector makes with a fixed direction in the plane. The bases of the prism correspond to each other in the manner determined by a translation of one into the other. The edges perpendicular to the bases all represent the same closed curve in the manifold of states of motion.

To see how the lateral faces correspond, observe that a vector on an edge of the plane polygon making an angle ψ with the outward-drawn normal corresponds to a vector at the corresponding point of the corresponding edge which makes with the inward-drawn normal an angle $-\psi$ or ψ according as the corresponding edges in question are described in the same or opposite senses in going around the polygon. From this it follows by means of elementary plane geometry that the inclination φ' of a vector on the edge a_i^{-1} is related to the inclination φ of the corresponding vector on a_i, both vectors being at points having angular distance θ from the initial points of their respective edges, by the formula

$$\varphi' = \varphi + 2\theta - 2m; \tag{2}$$

and similarly for the sides b_i and b_i^{-1}. For a pair of adjacent corresponding edges d_i the relation is

$$\varphi' = -\varphi - 2\theta + \frac{\pi}{k} + 2\varphi_0, \tag{3}$$

φ_0 being a constant representing the inclination of the outward-drawn normal to the first of the two sides d_i when $\theta = 0$.

On the vertical edges of the fluted prism select a set of points $A_1, A_2, \cdots, A_{2k}$, all representing the same point A of the manifold, such that A_1 is on the

edge passing through the initial point of a_1, A_2 is on the edge passing through the terminal point of a_1, and the other points lie successively, one on each edge, in the order thus determined. Let α_i be an arbitrary directed curve running from A_{4i-3} to A_{4i-2} and lying on that lateral face of the prism which is bounded in part by a_i. Let β_i be an arbitrary curve running from A_{4i-2} to A_{4i-1} and lying on that lateral face of the prism which is bounded in part by b_i. Let γ be the curve in the manifold beginning and ending at A which is represented by the vertical edges of the prism.

On that lateral face of the prism which is bounded in part by a_i^{-1} lies a curve α_i' representing the same curve in the manifold as α_i. Recalling that θ varies along each side between 0 and m, it will be seen that (2) shows that in going along α_i from A_{4i-3} to A_{4i-2} and then along α_i' from A_{4i-1} to A_{4i} the sum of the variations of φ is $-2m$. Similarly the sum of

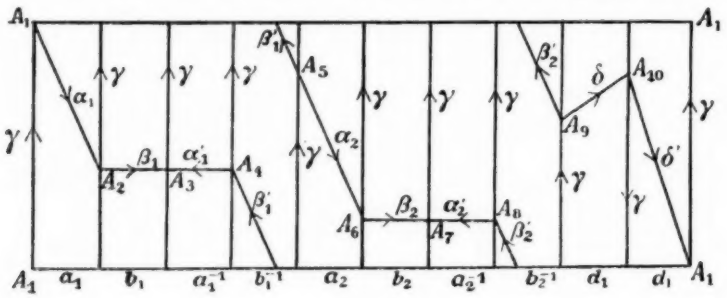


Fig. 6

the variations of φ in going along β_i from A_{4i-2} to A_{4i-1} and then along the corresponding curve β_i' from A_{4i} to A_{4i+1} is $-2m$. Thus in passing continuously around the prism the total variation of φ for each group of faces incident with the group of edges $a_i b_i a_i^{-1} b_i^{-1}$ is $-4m$.

For a pair of lateral faces incident with a pair of corresponding edges d_i the total variation of φ in going along a curve δ_i on the first face and then along the corresponding curve δ_i' on the second face is, according to (3), $-2m$.

For an orientable surface there are $p = k/2$ groups of sides $a_i b_i a_i^{-1} b_i^{-1}$ and no d's; the total variation of φ in going around the prism is thus $-4mp = -4\pi(1-2/k)\cdot k/2 = -2\pi(k-2)$.

For a one-sided surface of odd connectivity, $p = (k-1)/2$ and there is one pair of d's, so that the variation of φ is $-4m(k-1)/2 - 2m = -2\pi(k-2)$.

For a one-sided surface of even connectivity, $p = (k-2)/2$ and there are two pairs of d's, so that φ varies through $-4m(k-2)/2 - 4m = -2\pi(k-2)$.

A map of the lateral faces of the prism for the case $k = 5$, developed upon a plane, is shown in Fig. 6.

To obtain the Heegaard diagram, which is of genus $k+1$, each of the curves $\alpha_i, \beta_i, \gamma, \delta_i$ in the manifold is surrounded by a fine tube beginning and ending in a ball surrounding A. The $k+1$ curves which bound in the remainder of the manifold are those which bound a base of the prism and the k pairs of faces. In this way we find that the group has $k+1$ generators connected by the following $k+1$ generating relations:

For the orientable case,

$$\begin{aligned} &\gamma^{k-2}\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_p\beta_p\alpha_p^{-1}\beta_p^{-1} = 1, \\ &\alpha_i\gamma\alpha_i^{-1}\gamma^{-1} = 1, \\ &\beta_i\gamma\beta_i^{-1}\gamma^{-1} = 1 \qquad \left(i = 1, 2, \cdots, p; p = \frac{k}{2}\right). \end{aligned}$$

For the one-sided case, k odd,

$$\begin{aligned} &\gamma^{k-2}\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_p\beta_p\alpha_p^{-1}\beta_p^{-1}\delta^2 = 1, \\ &\alpha_i\gamma\alpha_i^{-1}\gamma^{-1} = 1, \\ &\beta_i\gamma\beta_i^{-1}\gamma^{-1} = 1, \\ &\gamma\delta\gamma\delta^{-1} = 1 \qquad \left(i = 1, 2, \cdots, p; p = \frac{k-1}{2}\right). \end{aligned}$$

For the one-sided case, k even,

$$\begin{aligned} &\gamma^{k-2}\alpha_1\beta_1\alpha_1^{-1}\beta_1^{-1}\cdots\alpha_p\beta_p\alpha_p^{-1}\beta_p^{-1}\delta_1^2\delta_2^2 = 1, \\ &\alpha_i\gamma\alpha_i^{-1}\gamma^{-1} = 1, \\ &\beta_i\gamma\beta_i^{-1}\gamma^{-1} = 1, \\ &\gamma\delta_1\gamma\delta_1^{-1} = 1, \\ &\gamma\delta_2\gamma\delta_2^{-1} = 1 \qquad \left(i = 1, 2, \cdots, p; p = \frac{k-2}{2}\right). \end{aligned}$$

From the groups the connectivities and coefficients of torsion are easily found.* The connectivity of the manifold of states of motion on a closed surface is in every case greater by unity than the connectivity of the surface. If the surface is orientable there is one coefficient of torsion, equal to $k-2$. If the surface is non-orientable and k is even, there are

* Veblen, loc. cit., Chap. V.

two coefficients of torsion, each equal to 2; if k is odd, 4 is the only coefficient of torsion. In every case the 3-dimensional manifold is orientable.

The results of the present section are true for all values of k. If $k = 2$ the argument requires only slight verbal modifications made necessary by the fact that the sides of the polygon are straight, forming a rectangle. The formulas pertaining to the groups agree with the results found in §§ 3 and 5 for $k = 0$ and $k = 1$, respectively.

11. If opposite directions are considered identical the groups described in § 10 are changed only by the substitution of γ^2 for γ in the first of the generating relations. The connectivities are unchanged. The coefficients of torsion are $2k-4$ if the surface is orientable, 2 and 2 if it is one-sided. These results, for $k = 0$ or 1, agree with those already found.

The Riemann spaces described in § 9 are the manifolds for the surfaces of connectivity 3 in this case also.

12. **Another method.** The groups and Heegaard diagrams found in § 10 for orientable surfaces can also be found by a generalization of the method of § 3. For an orientable surface of genus p the manifold of states of motion can be represented by a pair of regions, each of which is bounded by $p+1$ coaxial tores, one enclosing all the others. Since a Heegaard diagram must have only one surface for each region, and must be representable in a non-singular manner in euclidean space, we connect the tores bounding each region by p holes, transferring the 3-cells removed to the other region, where they appear as handles running through the holes. The regions thus obtained are canonical and of genus $2p+1$, since they are made simply connected by the following $2p+1$ non-intersecting cuts: (1) a cut whose bounding curve runs along a generating circle of each tore; (2) a cut severing each of the p handles; (3) a cut which enlarges each of the p holes so as to extend all the way around the common axis of the tores. By tracing upon the surfaces of one region the curves which bound in the other region (determined as in § 3) the Heegaard diagram and group are obtained.

13. **Boundaries.** The field of force for a dynamical problem may be such that there exist on the characteristic surface points where the velocity must vanish. A continuous closed curve consisting of such points is called by Birkhoff (loc. cit.) an oval of zero velocity.

An orbit which meets an oval of zero velocity must be normal to it, since a particle starting from rest must move along a curve tangent to the line of force through the initial point, and since the lines of force are normal to the oval of zero velocity, as well as to other equipotentials.*

* This is obvious in the reversible case. The following analytic proof, for which I am indebted to Professor C. E. Hille, holds in general. *Contin. p. 341.*

Since there is only one possible direction of motion at each point, the part of the manifold of states of motion corresponding to an oval of zero velocity is a curve, though the part corresponding to any other curve is a surface.

An isolated point P of zero velocity implies that the states of motion cannot be represented by a true manifold of three dimensions, since every neighborhood of the point corresponding to P represents the states of motion on a family of closed curves such as concentric circles, and so must be homeomorphic with a family of coaxial anchor rings, one enclosing all the others. The states of motion will constitute a *generalized manifold* of the type described by Dehn and Heegaard.* Any singularity of an oval

---

The differential equations of the motion, taken in Birkhoff's normal form

$$x'' + \lambda(x, y)y' = \frac{\partial}{\partial x} g(x, y), \qquad y'' - \lambda(x, y)x' = \frac{\partial}{\partial y} g(x, y),$$

where primes denote differentiation with regard to t and isothermal coördinates are used, reduce by the substitution

$$x_1 = x, \qquad x_2 = y, \qquad x_3 = x', \qquad x_4 = y'$$

to the system of first-order equations

$$x_1' = x_3, \qquad x_3' = -\lambda(x_1, x_2)x_4 + g_1(x_1, x_2),$$
$$x_2' = x_4, \qquad x_4' = \lambda(x_1, x_2)x_3 + g_2(x_1, x_2).$$

From the general theory of differential equations it follows that x_1, x_2, x_3 and x_4 can be expressed as power series in t which are convergent for sufficiently small values of t and which satisfy the differential equations. Mapping the characteristic surface conformally upon the xy-plane and taking the oval of zero velocity $g = 0$ through the origin, we may write, if $x_1 = x_2 = x_3 = x_4 = 0$ when $t = 0$,

$$g = ax + by + \cdots,$$
(i) $$x = \alpha_2 t^2 + \cdots,$$
(ii) $$y = \beta_2 t^2 + \cdots,$$

where the terms neglected are of higher orders. Substituting in the differential equations and equating constant terms we see that $2\alpha_2 = a$ and $2\beta_2 = b$. The slope of the tangent to the orbit at the origin is $\beta_2/\alpha_2 = b/a$. But the slope of $g = 0$ is $-a/b$. Hence the two lines are perpendicular.

Further, the orbit has a cusp at the origin. For by means of (i) t may be expressed as a power series in $\sqrt{x}$, and by means of (ii) t may be expressed as a power series in $\sqrt{y}$. Equating these two series and dropping higher order terms yields the equation of a semicubical parabola. It is of course possible that the time may become infinite as the particle approaches the origin. In this case the above argument requires only slight modification.

*In the Encyclopädie article on analysis situs; or cf. Veblen, loc. cit., p. 92.

of zero velocity, or an oval of zero velocity which does not form a true boundary between a region of positive velocity and a region not of positive velocity, will be represented in the manifold of states of motion by a singularity. We shall hereafter consider only non-singular ovals of zero velocity which are true boundaries of the characteristic surface. Since the stream lines are to be free from singularities we shall stick to the case in which opposite directions are considered distinct.

A short segment having one end on such a boundary will be represented in the manifold of states of motion by the interior of a circle, the center representing the point on the boundary. If the segment moves along the boundary and sweeps out a band, the circle will move through the manifold and sweep out a tore. Let us call this tore T_1. Variation through 2π of direction at a point of the surface is represented in the manifold by a curve which is deformable into a generating circle of this tore and hence to a point. An orbit having a cusp on the oval of zero velocity is represented in the manifold by a smooth curve through the center of one of these generating circles, with no singularity.

We are now in a position to describe the three-dimensional manifold of the states of motion on a characteristic surface bounded by a number ovals of zero velocity. The characteristic surface may first be closed by filling in a number of 2-cells bounded by the ovals. The manifold of possible states of motion on such a closed surface is studied in § 10 and we shall call this manifold M. Now let us remove a 2-cell containing one of the ovals.

Referring to § 10, it will be seen that removal of a 2-cell from the characteristic surface is equivalent to removal from the prism there used to represent the manifold M of a cylinder with generators parallel to those of the prism. As the ends of the prism are supposed to be joined, this cylinder represents a tore, T_1', in the manifold. Its removal means that the first generating relation given in § 10 for the group of each manifold no longer holds good, since a curve surrounding a base of the prism no longer bounds. Let M_1' be the open manifold obtained from M by removing the tore T_1' and let M_1 be the manifold obtained by putting T_1 back in place of T_1'.* The curves γ on the surface of T_1' are the boundaries of a moving 2-cell which generates T_1. In the group, γ now becomes the identity. Hence the generating relations given in § 10 which are not false when applied to M_1 are redundant.

The group of the manifold for a surface of connectivity $k+1$ having one oval of zero velocity thus has k entirely independent generators.

* We may describe these operations in other words as removing the interior of a tore and putting it back with interchange of bounding and non-bounding curves. The resulting figure cannot lie in euclidean 3-space.

Consider now the effect upon the group of the manifold of introducing a second non-singular oval of zero velocity. Performing a sequence of operations similar to that described above, we first form an open manifold M_2' by removing from M_1 a tore T_2', and then replace T_2' by a tore T_2 analogous to T_1. Let M_2 be the resulting manifold. Let α_{k+1} be the curve of intersection of the base $\varphi = 0$ of the prism with the surface of T_2'. Then α_{k+1} bounds a 2-cell in M_1 and T_2' but not in M_2', T_2 or M_2.

The group of M_2 has as generators α_{k+1} and the generators $\alpha_1, \cdots, \alpha_k$ of the group of M_1, and contains no generating relation. For since the group of M_1 contains no generating relation, the opposite supposition would be equivalent to assuming that a curve C exists which bounds a 2-cell in T_2 but not in M_2'. But it is one of the elementary properties of the tore in euclidean 3-space that a curve which bounds a 2-cell in its interior must be a curve of the class of a generating circle—counted perhaps more than once. Therefore, since C bounds a 2-cell in T_2, C is a power of γ. Since γ bounds a 2-cell in M_2' and C by assumption does not, we have a contradiction.

By repetition of this argument it may be shown that introduction of new non-singular ovals of zero velocity has no other effect on the group than to introduce new and independent generators. Our general result, therefore, is that if m ovals of zero velocity are introduced into a surface of connectivity $k+1$ the group of the manifold of states of motion has $k+m-1$ generators and no generating relation. The connectivity is therefore $k+m$.

14. We now show that this manifold can be represented by a two-sheeted Riemann space having $k+m$ non-linking and knotless branch curves.

The group of a surface of connectivity $k+1$ from which m 2-cells are removed has $n = k+m-1$ independent generators. Let A be a point in which n independent closed curves intersect. Let each of the removed 2-cells be enlarged until nothing is left of the surface but a set of n narrow bands along each of which one of the generating curves runs and a simply connected region about A which adjoins both ends of each band.

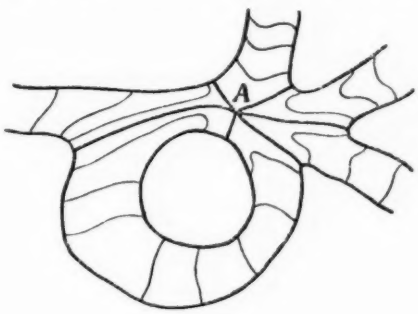


Fig. 7

For each generator g_i let there be a one-parameter continuous family C_i $(i = 1, 2, \cdots, n)$ of non-intersecting open curves on the surface having both ends on the boundary, each curve intersecting g_i once, and the first and

last curves of C_i passing through A (Fig. 7). The boundaries being ovals of zero velocity, the part of the manifold of states of motion corresponding to each curve of C_i is topologically a sphere, since it can be described by a one-parameter family of circles of which the first and last are of zero radius, each circle representing the possible directions at one point of the curve. The manifold therefore consists of n one-parameter families of spheres, half of each end sphere being identified with half of some other end sphere.

Each of these families of spheres may be regarded as lying in and filling the portion of a space of inversion between two spheres. This region is homeomorphic with a 2-sheeted Riemann space from each sheet of which a sphere has been removed, the branch system being a single circle.

The manifold is therefore homeomorphic with a Riemann space of $2n$ sheets with one circular branch curve about which all the sheets permute in the same cyclic order as that in which the corresponding ends of the bands are arranged about A, and n other non-linking circular branch curves connecting the sheets in pairs in a manner corresponding to that in which the bands connect their ends.

By a reduction similar to that of Riemann surfaces by Clebsch and Lüroth this Riemann space may be shown to be homeomorphic with a two-sheeted space with non-linking and knotless branch curves. These curves must be $k+m$ in number because it was proved in § 13 that the connectivity of the manifold is $k+m$. This value of the connectivity may also be inferred from the present section without reference to § 13.

15. **Summary.** We have first found the simple manifolds of states of motion for the sphere and projective plane and then obtained general formulas for the groups and have shown how to construct the Heegaard diagrams for closed surfaces in general. The general method used applies with slight modifications to the sphere and projective plane. For surfaces of the connectivity of the anchor ring, Riemann spaces of simple type homeomorphic with the manifolds of states of motion were found.

Open surfaces, which appear in dynamical problems in which the potential functions are such as to delimit regions which the moving particle cannot enter, were considered in §§ 13 and 14 from different points of view. In § 13 the groups were deduced from the results already found for closed surfaces. In § 14 a method of attack yielding extremely simple Riemann spaces was used.

The writer wishes to acknowledge his indebtedness to Professor J. W. Alexander for many suggestions.

Stanford University, Calif.

RELATIONS BETWEEN THE CRITICAL POINTS OF A REAL FUNCTION OF n INDEPENDENT VARIABLES*

BY

MARSTON MORSE

INTRODUCTION

In the main body of the paper the results and proofs are given for the general case of a real function of n independent variables, and for a domain of definition of the given function that is very general from the point of view of analysis situs. For the purposes of an introduction one of the principal results will be given for the case where $n = 2$. It must be stated, however, that the results for the general case differ tremendously from those for the case $n = 2$. The results for the case $n \leq 2$ can conveniently be given in geometric language as follows:

Let R consist of the interior points of a finite, connected region of the x, y plane, bounded by a finite number of distinct closed curves without multiple points of any sort, possessing at each point a tangent which turns continuously with movement of its point of contact. Denote the boundary of R by B. Let $f(x, y)$ be a real function of x and y defined and continuous in R and on B, and possessing continuous third order partial derivatives in R taking on continuous boundary values on B. By the positive direction of the normal to B will be understood the direction that leads across B from points in R to points not in R. It is assumed that the unilateral, directional derivative of $f(x, y)$ on the side of the inner half of the normal and in the positive direction along the normal is everywhere positive on B.

Let $z = f(x, y)$ be interpreted as a surface in ordinary euclidean (x, y, z) space. By a critical point of $f(x, y)$ is meant a point in the x, y plane corresponding to which the tangent plane to the surface $z = f(x, y)$ is parallel to the x, y plane. In the neighborhood of such a critical point let $f(x, y)$ be expanded according to Taylor's formula in powers of differences of the x's and y's which vanish at the critical point, and with the remainder as a term of the third order. We hereby assume that in this expansion the discriminant of the terms of the second order is not zero.

From these assumptions it follows that the critical points are isolated and finite in number, and that they are of the two general types called

* Presented to the Society, December 28, 1923.

in differential geometry elliptic and hyperbolic points. The elliptic critical points can be further subdivided into points at which $f(x, y)$ takes on a relative minimum, and those at which $f(x, y)$ takes on a relative maximum. Suppose that $f(x, y)$ has a relative minimum at M_0 different points, and a relative maximum at M_2 different points, while the hyperbolic critical points are in number M_1. Suppose that R and its boundary form a region of linear connectivity R_1.* With this understood we have the equality fundamental for the case $n = 2$,

$$M_0 - M_1 + M_2 = 2 - R_1 . \tag{a}$$

From this it follows immediately that

$$M_1 \geqq M_0 + R_1 - 2 . \tag{b}$$

It should be noted that (a) does not follow from (b). The inequality (b) has been obtained by G. D. Birkhoff.† Enunciated in slightly different form for the general case, he has called his result the "minimax principle" and he uses it to infer the existence of at least $M_0 + R_1 - 2$ critical points of the hyperbolic type when the existence of M_0 points of relative minimum are known. He has carried out his work for the general case of real, analytic function of n real variables and applied his result to the theory of dynamical systems. Upon reading Birkhoff's paper it occurred to the author that inasmuch as there are $n+1$ different kinds of critical points possible (in a sense to be defined later) there ought to be relations analogous to (b) involving the numbers of critical points of all kinds and not simply of just two kinds. With a proper interpretation of M_1 and R_1, (b) however will serve to state Birkhoff's "minimax principle" in any number of dimensions.

The author has replaced the inequality (b) by an equality involving all of the $n+1$ possible types of critical points, and in addition all of the connectivity numbers of the n-dimensional region of definition of the given function (Theorems 8 and 9). Further, a set of n additional inequality relations have been discovered. The author ventures without proof the opinion that these relations, or relations derivable from them, are all the the relations there are between the numbers involved.

* Oswald Veblen, *Analysis Situs*, *The Cambridge Colloquium*, Part 2, p. 50, § 28.

We shall have occasion in the future as well as here to make use of the terms of analysis situs. They will be used in the senses in which they are defined in the work just cited. In the remainder of this paper reference to this work will be indicated by the letter V.

† *Dynamical systems with two degrees of freedom*, these Transactions, vol. 18 (1917), p. 240.

Note. For the case $n = 2$ the author has been able to outline an alternative proof of one of the relations between the critical points by making use of the Kronecker characteristic of the given function and its partial derivatives. That the Kronecker theory as developed by Kronecker (*Werke*, vol. I, pp. 175–226, and vol. II, pp. 71–82) will not suffice in general to obtain the results of this paper is evident from the fact that Kronecker distinguishes between the critical points according to the sign of the hessian of the function, while in this paper both in the proofs and in the results there is an essential distinction between the $n+1$ types of critical points. Further, in Kronecker no reference is made to anything like the connectivity numbers of the regions in which the critical points are being studied.*

THE FUNCTION $f(x_1, x_2, \cdots, x_n)$

1. The function $f(x_1, x_2, \cdots, x_n)$ and its critical points. Let $(x_1, x_2, \cdots, x_n)$ be the coördinates of any point (X) in a euclidean space of n dimensions. In a finite part of this space let there be given an n-dimensional region Σ. The set of interior points of Σ will be denoted by R. At every point of R let there be defined a real, single-valued, continuous function of the x's, namely $f(x_1, x_2, \cdots, x_n)$. Suppose further that in R this function possesses continuous partial derivatives with respect to each of the x's of at least the first and second orders.

A point of R at which all of the first partial derivatives of f vanish will be called a *critical point* of f. Any other point of R will be called an *ordinary point* of f. The value of f at a critical point will be termed a *critical value* of f. As a further assumption regarding f we will suppose that in the neighborhood of each critical point f possesses continuous third partial derivatives. Let there be given a critical point of f at which f takes on the value c. For simplicity suppose this critical point at the origin. In the neighborhood of the origin, f can be represented with the aid of Taylor's formula in the form

$$f(x_1, x_2, \cdots, x_n) - c = \sum_{ij} a_{ij} x_i x_j + r(x_1, x_2, \cdots, x_n) \tag{1}$$

$$(i, j = 1, 2, \cdots, n; \quad a_{ij} = a_{ji}),$$

where the terms of first order are missing because the origin is a critical point, where the terms of second order are represented as a symmetric quadratic form in the n x's with constant coefficients, and where the remainder $r(x_1, x_2, \cdots, x_n)$ is included as a term of the third order. In

* For closely connected results in the case $n = 2$ see a paper by H. Poincaré, Journal de Mathématiques, ser. 3, vol. 7 (1881), pp. 375–421.

any representation such as this of the function f in the neighborhood of a critical point, it is hereby assumed that the terms of the second order constitute a non-singular quadratic form. Analytically this means that the determinant

$$|a_{ij}| \neq 0. \tag{2}$$

It is a consequence of this last restriction upon f that *in a sufficiently small neighborhood of any critical point there is no other critical point.* For, if the partial derivative of f with respect to x_i be denoted by f_i $(i = 1, 2, \cdots, n)$, each critical point of f is a solution of the equations

$$f_1 = f_2 = \cdots = f_n = 0. \tag{3}$$

Now the jacobian of the n functions f_i with respect to the n variables x_i at the critical point we are here considering, namely the origin, vanishes only with the determinant $|a_{ij}|$ which we have assumed not zero. It follows from the theory of implicit functions that the system of equations (3) has no other solution in the neighborhood of the origin than the origin itself. Thus there are no other critical points in the neighborhood of the origin.

A consequence of the result just obtained is that *in any closed region S all of whose points are points of R, there are at most a finite number of critical points of the given function.* For if there were an infinite number of critical points in S, by the Weierstrass cluster point theorem, these critical points would have at least one cluster point, which, since S is closed, would belong to S, and hence would be a point of R. At each of these critical points all of the first partial derivatives of f would vanish, and since these partial derivatives are assumed continuous at every point in R, they would all be zero at the cluster point. Thus the cluster point would also be a critical point contrary to the result of the preceding paragraph that the critical points of f are isolated.

2. **Certain preliminary lemmas.** Lemmas 1 and 2 of this section are important aids in later proofs. They refer to a function $F(x_1, x_2, \cdots, x_n)$, real, single-valued, and continuous in the neighborhood of the origin (written N). To aid in proving Lemma 1, results (A), (B), and (C) will first be established.

(A) *If $F(x_1, x_2, \cdots, x_n)$ possesses continuous first partial derivatives in N, then in N*

$$F(x_1, x_2, \cdots, x_n) = F(0, x_2, x_3, \cdots, x_n) + a(x_1, x_2, \cdots, x_n)\, x_1 \tag{4}$$

where $a(x_1, x_2, \cdots, x_n)$ is single-valued and continuous in N.

A function $a(x_1, x_2, \cdots, x_n)$ readily seen to satisfy the requirements of (A) is defined by the equations

$$(5) \qquad a(x_1, x_2, \cdots, x_n) = \frac{F(x_1, x_2, \cdots, x_n) - F(0, x_2, \cdots, x_n)}{x_1}, \quad x_1 \neq 0,$$

$$(6) \qquad a(0, x_2, x_3, \cdots, x_n) = F_1(0, x_2, \cdots, x_n),$$

where F_i will be taken to mean the partial derivative of F with respect to x_i $(i = 1, 2, \cdots, n)$.

(B) *If $F(x_1, x_2, \cdots, x_n)$ possesses continuous second partial derivatives in N, then in N*

$$(7) \qquad F(x_1, x_2, \cdots, x_n) = F(0, x_2, \cdots, x_n) + F_1(0, x_2, \cdots, x_n)\, x_1 + b(x_1, x_2, \cdots, x_n)\, x_1^2$$

where $b(x_1, x_2, \cdots, x_n)$ is single-valued and continuous in N.

A function $b(x_1, x_2, \cdots, x_n)$ readily seen to satisfy the requirements of (B) is defined by the equations

$$(8) \quad b(x_1, x_2, \cdots, x_n) = \frac{F(x_1, x_2, \cdots, x_n) - F(0, x_2, \cdots, x_n) - F_1(0, x_2, \cdots, x_n)x_1}{x_1^2}, \quad x_1 \neq 0,$$

$$(9) \quad b(0, x_2, \cdots, x_n) = \frac{F_{11}(0, x_2, \cdots, x_n)}{2},$$

where F_{ij} will be taken to mean the second partial derivative of F with respect to x_i and x_j $(i, j = 1, 2, \cdots, n)$.

(C) *If $F(x_1, x_2, \cdots, x_n)$ possesses continuous first partial derivatives in N, then in N*

$$(10) \quad F(x_1, x_2, \cdots, x_n) = F(0, 0, \cdots, 0) + \sum_i a_i(x_1, x_2, \cdots, x_n)x_i \quad (i = 1, 2, \cdots, n),$$

where each a_i can be so chosen as to be continuous in N.

From (A) it follows that (C) is true for $n = 1$. Following the method of mathematical induction the truth of (C) is now assumed for the case of $n-1$ independent variables. Whence

$$(11) \qquad F(0, x_2, \cdots, x_n) = F(0, 0, \cdots, 0) + \sum_i b_i(x_2, \cdots, x_n)\, x_i \quad (i = 2, \cdots, n)$$

where each b_i is continuous in N. If $F(0, x_2, \cdots, x_n)$ in (4) be replaced by the right hand member of (11) the result (C) follows directly.

LEMMA 1. *If $F(x_1, x_2, \cdots, x_n)$ possesses continuous second partial derivatives in N, and vanishes with its first partial derivatives at the origin, then in N*

$$(12)\quad F(x_1, x_2, \cdots, x_n) = \sum_{ij} a_{ij}(x_1, x_2, \cdots, x_n)\, x_i x_j \qquad (i, j = 1, 2, \cdots, n)$$

where each function a_{ij} can be so chosen as to be continuous in N.

For $n = 1$ (B) shows that Lemma 1 is true. If Lemma 1 be assumed to be true for the case of $n-1$ independent variables, then

$$(13)\quad F(0, x_2, \cdots, x_n) = \sum_{ij} b_{ij}(x_2, \cdots, x_n)\, x_i x_j \qquad (i, j = 2, 3, \cdots, n)$$

where each b_{ij} is continuous in N. If $F(x_1, x_2, \cdots, x_n)$ satisfies the hypotheses of this lemma, $F_1(0, x_2, \cdots, x_n)$ will satisfy the hypotheses (C) so that

$$(14)\quad F_1(0, x_2, \cdots, x_n) = \sum_{i} c_i(x_2, \cdots, x_n)\, x_i \qquad (i = 2, 3, \cdots, n)$$

where each c_i is continuous in N. If now in (7) $F(0, x_2, \cdots, x_n)$ and $F_1(0, x_2, \cdots, x_n)$ be replaced by their values from (13) and (14) respectively, then Lemma 1 follows as stated.

To prove Lemma 2, the following results, (A′), (B′), and (C′), will first be proved.

(A′) *If F possesses continuous second partial derivatives in N then the function $a(x_1, x_2, \cdots, x_n)$ of* (A) *as defined by* (5) *and* (6) *has continuous first partial derivatives in N.*

It follows from (5) that $a(x_1, x_2, \cdots, x_n)$ has continuous first partial derivatives in N if $x_1 \neq 0$. For $x_1 = 0$ the partial derivatives of $a(x_1, x_2, \cdots, x_n)$ with respect to $(x_2, x_3, \cdots, x_n)$ exist, as follows from (6). For $x_1 \neq 0$ it follows from Taylor's formula that

$$(15)\quad a(x_1, x_2, \cdots, x_n) = F_1(0, x_2, \cdots, x_n) + \frac{F_{11}(\theta x_1, x_2, \cdots, x_n)\, x_1}{2}, \quad 0 < \theta < 1.$$

From (15) and (6) and the definition of a derivative it follows that the partial derivative

$$a_{x_1}(0, x_2, \cdots, x_n) = \frac{F_{11}(0, x_2, \cdots, x_n)}{2}.$$

To prove $a_{x_1}(x_1, x_2, \cdots, x_n)$ continuous for $x_1 = 0$, observe that from (5) for $x_1 \neq 0$ it follows that

$$a_{x_1} = \frac{x_1 F_1(x_1, x_2, \cdots, x_n) - [F(x_1, x_2, \cdots, x_n) - F(0, x_2, \cdots, x_n)]}{x_1^2},$$

which by two applications of Taylor's formula becomes

$$\frac{x_1[F_1(0, x_2, \cdots, x_n) + F_{11}(\theta x_1, x_2, \cdots, x_n)x_1]}{x_1^2} - \frac{F_1(0, x_2, \cdots, x_n)x_1 + \dfrac{F_{11}(\theta' x_1, x_2, \cdots, x_n)x_1^2}{2}}{x_1^2}$$

$$= F_{11}(\theta x_1, x_2, \cdots, x_n) - \frac{F_{11}(\theta' x_1, x_2, \cdots, x_n)}{2}, \quad 0 < \theta < 1, \ 0 < \theta' < 1,$$

whence

$$\lim_{x_1 \to 0} a_{x_1}(x_1, x_2, \cdots, x_n) = \frac{F_{11}(0, x_2, \cdots, x_n)}{2}$$

and a_{x_1} is thereby proved continuous in N without exception.

To show that $a_{x_2}(x_1, x_2, \cdots, x_n)$ is continuous in N for $x_1 = 0$ observe that from (5) for $x_1 \neq 0$ it follows that

$$a_{x_2}(x_1, x_2, \cdots, x_n) = \frac{F_2(x_1, x_2, \cdots, x_n) - F_2(0, x_2, \cdots, x_n)}{x_1},$$

which by Taylor's formula gives

(16) $\quad a_{x_2}(x_1, x_2, \cdots, x_n) = F_{21}(\theta x_1, x_2, \cdots, x_n), \qquad 0 < \theta < 1, \ x_1 \neq 0,$

while from (6)

(17) $\qquad a_{x_2}(0, x_2, \cdots, x_n) = F_{12}(0, x_2, \cdots, x_n).$

From (16) and (17) it follows that a_{x_2} is continuous in N even when $x_1 = 0$. The remaining partial derivatives of $a(x_1, x_2, \cdots, x_n)$ are similarly seen to be continuous in N.

(B′) *If F possesses continuous third partial derivatives in N, then the function $b(x_1, x_2, \cdots, x_n)$ of* (B) *as defined by* (8) *and* (9) *possesses continuous first partial derivatives in N.*

The proof of (B′) is similar to that of (A′).

(C′) *If F possesses continuous second partial derivatives in N then the functions $a_i(x_1, x_2, \cdots, x_n)$ of* (C) *can be so chosen as to possess continuous first partial derivatives in N.*

A review of the proof of (C) with (A′) in mind will show the truth of (C′).

LEMMA 2. *If F possesses continuous third partial derivatives in N, and vanishes with its first partial derivatives at the origin, then F can be represented in the form*

(18) $\qquad F(x_1, x_2, \cdots, x_n) = \sum_{ij} a_{ij} x_i x_j \quad (i, j = 1, 2, \cdots, n)$

where each a_{ij} possesses continuous first partial derivatives in N.

A review of the proof of Lemma 1 with (A′), (B′) and (C′) in mind will show the truth of Lemma 2.

LEMMA 3. *The functions* $a_{ij}(x_1, x_2, \cdots, x_n)$ *of Lemmas* 1 *and* 2 *are such that at the origin*

$$a_{ij} + a_{ji} = F_{ij}. \tag{19}$$

Under the hypotheses of Lemmas 1 and 2

$$F(x_1, x_2, \cdots, x_n) = \frac{\sum_{ij} F_{ij}(\theta_1 x_1, \theta_2 x_2, \cdots, \theta_n x_n)\, x_i x_j}{2}, \quad 0 < \theta_i < 1. \tag{20}$$

Hence from (20) and (18)

$$\sum_{ij} a_{ij} x_i x_j = \frac{\sum_{ij} F_{ij}(\theta_1 x_1, \theta_2 x_2, \cdots, \theta_n x_n)\, x_i x_j}{2}. \tag{21}$$

Let $(y_1, y_2, \cdots, y_n)$ be any point in N, and let r be a positive constant so small that $(ry_1, ry_2, \cdots, ry_n)$ is a point in N for any choice of $(y_1, y_2, \cdots, y_n)$ in N. In (21) set $x_i = ry_i$ and in the resulting equation cancel the factor r^2 from both members. There will then result the equation

$$\sum_{ij} a_{ij}(ry_1, ry_2, \cdots, ry_n)\, y_i y_j = \frac{\sum_{ij} F_{ij}(r\theta_1 y_1, r\theta_2 y_2, \cdots, r\theta_n y_n)\, y_i y_j}{2}. \tag{22}$$

The two members of (22) possess limits as r approaches zero, and these limits are equal, giving

$$\sum_{ij} a_{ij}(0, 0, \cdots, 0)\, y_i y_j = \frac{\sum_{ij} F_{ij}(0, 0, \cdots, 0)\, y_i y_j}{2}. \tag{23}$$

Equation (23) is an identity in the y's. Hence the coefficients of corresponding powers are equal. The equation (19) follows directly and the lemma is proved.

3. **Different types of critical points.** Let there be given a critical point of f. At this critical point let c be the value of f. If this critical point be supposed to lie at the origin, it follows from Lemma 2 that in the neighborhood of the origin the difference $f - c$ can be represented in the form

$$f(x_1, x_2, \cdots, x_n) - c = \sum_{ij} a_{ij}(x_1, x_2, \cdots, x_n)\, x_i x_j \quad (i, j = 1, 2, \cdots, n) \tag{24}$$

where $a_{ij}(x_1, x_2, \cdots, x_n)$ is a single-valued, continuous function of the x's in a sufficiently small neighborhood of the origin, provided with continuous

first partial derivatives with respect to each of the x's. Further without any loss of generality, we can suppose that

$$a_{ij} = a_{ji}. \tag{25}$$

If we set $a_{ij}(0, 0, \cdots, 0) = \alpha_{ij}$, according to Lemma 3, we will have at the origin

$$2\alpha_{ij} = f_{x_i x_j}. \tag{26}$$

Equation (26) shows that the constants α_{ij} appearing here are the same as the constants α_{ij} used in (1). In § 1 we assumed that $|\alpha_{ij}| \neq 0$. We can conclude that

$$|a_{ij}| \neq 0, \tag{27}$$

in a sufficiently small neighborhood of the origin.

According to the well known theory of quadratic forms a non-singular quadratic form in n variables $(x_1, x_2, \cdots, x_n)$, by a real, non-singular, linear transformation of $(x_1, x_2, \cdots, x_n)$ into n new variables $(y_1, y_2, \cdots, y_n)$, can be reduced to a quadratic form

$$-y_1^2 - y_2^2 - \cdots - y_k^2 + y_{k+1}^2 + \cdots + y_n^2 \tag{28}$$

where k is one of the integers from 0 to n, inclusive.* An examination of the Lagrange method of making this reduction shows that it can be carried through in the case where the coefficients a_{ij} are not constants, by formally following the method used for the case where the coefficients a_{ij} are constants, writing down the Lagrange transformation at each stage with the variable coefficients a_{ij} substituted for the constant coefficients a_{ij} of the ordinary theory, provided at each stage at least one of the coefficients a_{ii} is not zero at the origin. If at any stage all of the coefficients a_{ii} are zero at the origin, at least one of the remaining coefficients, say a_{rs}, will not be zero at the origin since $|\alpha_{ij}| \neq 0$. If the preliminary transformation

$$\begin{aligned} x_r &= z_r - z_s, \\ x_s &= z_r + z_s, \\ x_i &= z_i \qquad (i \neq r, s) \end{aligned}$$

be made the resulting coefficients of z_{rr} and z_{ss} will not be zero at the origin, and the Lagrange type of transformation can then be used. The transformation will of course be restricted to a neighborhood of the origin for which $|a_{ij}| \neq 0$ and such that each a_{ij} by which it has been necessary

* Bôcher, *Introduction to Higher Algebra*, p. 131.

to divide shall not be zero. If the given critical point is not at the origin a proper translation of the space of the n x's will bring the critical point to the origin. The above method of reduction of the function f will then give the following lemma.

LEMMA 4. *If $f(x_1, x_2, \cdots, x_n)$ takes on the value c at a critical point $(x_1^0, x_2^0, \cdots, x_n^0) = (X_0)$ then there exists a real transformation of the form*

$$y_i = \sum_j c_{ij}(x_j - x_j^0), \quad |c_{ij}| \neq 0 \quad (i, j = 1, 2, \cdots, n) \tag{29}$$

under which

$$f(x_1, x_2, \cdots, x_n) - c = -y_1^2 - y_2^2 - \cdots - y_k^2 + y_{k+1}^2 + \cdots + y_n^2 \tag{30}$$

where each c_{ij} is a continuous function of the x's provided with continuous first partial derivatives in a sufficiently small neighborhood of (X_0), and where k is some integer from 0 to n, inclusive.

The integer k is called the type of the given critical point. The number of critical points of the kth type under boundary conditions to be introduced presently will be shown to be finite and will be denoted by m_k. It is the purpose of this paper to obtain the relations between the numbers m_k under various types of boundary conditions.

The jacobian of the functions $(y_1, y_2, \cdots, y_n)$ of (29) with respect to the variables $(x_1, x_2, \cdots, x_n)$ reduces at (X_0) to the value of $|c_{ij}|$ at (X_0). Thus this jacobian is not zero at (X_0). It follows that *the transformation* (29) *is one-to-one in the neighborhood of the given critical point.*

4. **The boundary conditions α.** In order to proceed further it is necessary to make some assumptions regarding boundary conditions. The first boundary conditions that we introduce are, strictly speaking, not boundary conditions alone. They will be shown later to lead to results depending upon boundary conditions in a strict sense. These first boundary conditions which we consider are introduced because of their simplicity.

The boundary conditions α will be considered as fulfilled if among the $(n-1)$-dimensional spreads

$$f(x_1, x_2, \cdots, x_n) = \text{const.}$$

there exists one, say A, with the following properties:

I. *A is a closed manifold;*

II. *A forms the complete boundary of a portion of space, S, including A, all of whose points are points of R;*

III. *Each critical point of f is an interior point of S;*

IV. *The value of f on A is greater than the value of f at any interior point of S.*

The region S as defined in the boundary conditions α is a closed region containing all of the critical points of f. It follows from the results of § 1 that there are at most a finite number of critical points of f. According to condition III there is no critical point of f on A.

5. **The complex of points satisfying** $f(x_1, x_2, \cdots, x_n) \leqq c$. According to the boundary condition IV the absolute maximum of f in S is taken on at each point on the boundary of S and at no other point of S. If any absolute minimum point of f in S be joined by a continuous curve to some boundary point of S, every constant c between the absolute minimum m of f in S and the absolute maximum M of f in S will be taken on by f at some point in S on this curve. Let c be any such constant between m and M, not a critical value of f. The set of points $(x_1, x_2, \cdots, x_n) = (X)$, lying in S and satisfying

$$f(x_1, x_2, \cdots, x_n) = c \tag{31}$$

according to the choice of c, are all ordinary points of f. This set of points is closed since f is continuous. Its points are all interior points of S since the value of f on S's boundary is M. If (X_0) be any point on (31), at least one of the partial derivatives of f, say f_{x_1}, will not be zero at (X_0). It follows that the points $(x_1, x_2, \cdots, x_n)$ which satisfy (31) in the neighborhood of (X_0) can be represented in the form

$$x_1 = h(x_2, x_3, \cdots, x_n)$$

where $h(x_2, x_3, \cdots, x_n)$ is a single-valued function of its arguments provided with continuous partial derivatives up to the second order all in the neighborhood of (X_0). Geometrically this means that the manifold of points satisfying (31) possesses at each point of (31) a hypertangent plane which turns continuously as the point of tangency varies continuously on (31). The manifold (31) thus has no singular points.

The set of points (X) in S satisfying

$$f \leqq c \tag{32}$$

will make up an n-dimensional region, say C_n. In particular C_n will include all of the points neighboring any absolute minimum of f. Its points are all interior points of S, and its boundary obviously consists of the points (X) satisfying (31). It can be shown that C_n is an n-dimensional complex in the technical sense of analysis situs.* In a proof of this the same sort of considerations would enter as enter in connection with the so called "regular regions" in the plane.†

* See V, loc. cit., Chap. 3.

† W. F. Osgood, *Lehrbuch der Funktionentheorie*, p. 179.

The variation of the connectivity numbers $R_0, R_1, \cdots, R_n$ of C_n with variations of c will now be studied. To that end it will be convenient to consider first the trajectories orthogonal to the manifolds $f = \text{const.}$

6. **The trajectories orthogonal to the manifolds $f = \text{const.}$** In studying the trajectories orthogonal to the manifolds $f = c$ the neighborhoods of critical points of f will be excluded from consideration. The trajectories orthogonal to the manifolds $f = c$ at ordinary points of f might be defined as those curves along which the differentials $dx_1, dx_2, \cdots, dx_n$ at any point are proportional to the partial derivatives $f_1, f_2, \cdots, f_n$ at that point, where f_i is used to denote the partial derivative of f with respect to x_i $(i = 1, 2, \cdots, n)$. *Two trajectories are considered as the same* if they consist of the same points $(x_1, x_2, \cdots, x_n)$. It is thus explicitly pointed out that a difference of parametric representative alone does not, according to this convention, constitute a difference in the trajectories.

At any ordinary point of f at least one of the partial derivatives of f is not zero. Let (P) be an ordinary point of f and suppose that at (P) the partial derivative $f_m \neq 0$. The trajectories passing through the points neighboring (P) will be solutions of the differential equations

$$\frac{dx_i}{dx_m} = \frac{f_i}{f_m}, \tag{33}$$

where i takes on all the integers from 1 to n except m. According to the fundamental existence theorems for ordinary differential equations there is one and only one trajectory through each point of a sufficiently small neighborhood of (P).

A set of differential equations will now be introduced in which the independent variable is not x_m but a parameter τ, but whose solutions also represent orthogonal trajectories. The equations are

$$\frac{dx_i}{d\tau} = rf_i \qquad (i = 1, 2, \cdots, n) \tag{34}$$

where

$$r = \frac{1}{f_1^2 + f_2^2 + \cdots + f_n^2}.$$

The particular form of dependence of the x's upon τ which is required by the above differential equations (34) is chosen in preference to any other because for a solution $x_1(\tau), x_2(\tau), \cdots, x_n(\tau)$ of (34) the following identity in τ holds:

$$f[x_1(\tau), x_2(\tau), \cdots, x_n(\tau)] \equiv \tau + \text{const.} \tag{35}$$

For if the left hand member of (35) be differentiated with respect to r, making use of (34) it is found that the resulting derivative is unity. Whence (35) follows directly.

Let there be given an ordinary point (B) with coördinates $(b_1, b_2, \cdots, b_n)$. In a sufficiently small neighborhood of (B) the denominator of the quotient in (34) giving r is not zero, and in such a neighborhood the right hand members of the differential equations (34) possess continuous partial derivatives with respect to each of the x's. Further these right hand members do not involve t at all. It follows from the fundamental existence theorems of ordinary differential equations that there exists a positive constant e so small that for any point $(A) = (a_1, a_2, \cdots, a_n)$ in the e neighborhood of (B) and for any constant t_0 whatsoever, and range of t within e of t_0, there exist functions

$$x_i = h_i(a_1, a_2, \cdots, a_n, t_0, t) \qquad (i = 1, 2, \cdots, n) \tag{35'}$$

which give for constant values of $(a_1, a_2, \cdots, a_n, t_0)$ and for a variation of t, a solution of (34) that passes through (A) when $t = t_0$, and is the only such solution of (34). Further, for the above prescribed domain of $(a_1, a_2, \cdots, a_n, t_0)$ the right hand members of (35′) possess continuous partial derivatives of at least the first order with respect to all of their arguments.

Because the right hand members of (34) do not involve t it follows that two solutions of (34), for which the points $(a_1, a_2, \cdots, a_n)$ are the same but for which t_0 differs, will give the same set of points $(x_1, x_2, \cdots, x_n)$, assigned however by (35′) to different values of t. What is here desired is to obtain all of the orthogonal trajectories in a form which will be the simplest possible for our purposes. This end will be served by setting

$$t_0 = f(a_1, a_2, \cdots, a_n).$$

Upon writing

$$h_i[a_1, a_2, \cdots, a_n, f(a_1, a_2, \cdots, a_n), t] = H_i(a_1, a_2, \cdots, a_n, t) \; (i = 1, 2, \cdots, n)$$

(35′) reduces to

$$x_i = H_i(a_1, a_2, \cdots, a_n, t). \tag{36}$$

For initial conditions we now have

$$a_i = H_i[a_1, a_2, \cdots, a_n, f(a_1, a_2, \cdots, a_n)], \tag{37}$$

an identity in all of the a_i's in the neighborhood of (B). If the solution of (34), namely $x_1(t), x_2(t), \cdots, x_n(t)$, appearing in (35) be now understood

to be a solution given by (36), a substitution in (35) of $\tau = f(a_1, a_2, \cdots, a_n)$ will show that the constant in (35) is zero. Thus for all solutions of (34) given by (36) there results the identity in t

$$f[x_1(t), x_2(t), \cdots, x_n(t)] \equiv \tau. \tag{38}$$

Thus we have proved the following lemma:

LEMMA 5. *The trajectories orthogonal to the manifolds $f = c$ can be so represented that the parameter τ at any one of their points equals the value of f at that point.*

7. **Invariance of the connectivity numbers of the complex $f \leqq c$ with variation of c through ordinary values of f.** The following lemma will now be proved.

LEMMA 6. *Let a and b $(a < b)$ be two ordinary values of f in S such that the interval between a and b contains only ordinary values of f; then the set of points in S satisfying*

$$f \leqq a \tag{39}$$

can be put into one-to-one continuous correspondence with the set of points in S satisfying

$$f \leqq b. \tag{40}$$

The constant a is by hypothesis not a critical value of f. Since there are altogether at most a finite number of critical values of f there can be chosen a constant α less than a and differing from a by so little that between an a and α there is no constant equal to a critical value of f. Denote by H a set of orthogonal trajectories in which there is just one trajectory passing through each point of

$$f = \alpha. \tag{41}$$

Let these trajectories be represented in the form described in Lemma 5 in which the value of the parameter τ at each point equals the value of f at that point. The point on each trajectory that lies on (41) will correspond to the parameter value $\tau = \alpha$.

Since there is no critical value of f equal to a constant between α and b each trajectory that starts from a point of (41) can be continued through an arc corresponding to a range of values of τ from α to b. On these arcs the parameter values $\tau = a$ and $\tau = b$, respectively, will correspond to points on $f = a$ and $f = b$. To the complete set of values of τ satisfying

$$\alpha \leqq \tau \leqq b \tag{42}$$

there will correspond points satisfying

$$\alpha \leqq f \leqq b. \tag{43}$$

Further, the set of trajectories H, including as it does a trajectory through each point of (41), if continued through the interval (42) will include at least one trajectory through each point $(x_1, x_2, \cdots, x_n)$ satisfying (43). For if

$$(a_1, a_2, \cdots, a_n) = (A)$$

be any point such that

$$\alpha \leqq f(a_1, a_2, \cdots, a_n) \leqq b \tag{44}$$

a trajectory through (A), represented as described in Lemma 5, will pass through (A) for a parameter value $\tau = f(a_1, a_2, \cdots, a_n)$ and if continued in the sense of decreasing τ will reach a point on (41) when τ reaches the value $\tau = \alpha$. Finally, it follows from the fundamental existence theorem stated in § 6 that there is not more than one of these trajectories through each point satisfying (43).

The correspondence whose existence is affirmed in the lemma can now be set up. First observe that the points satisfying (39) consist of the sum of the points satisfying

$$f \leqq \alpha \tag{45}$$

and those satisfying

$$\alpha < f \leqq a, \tag{46}$$

while the points satisfying (40) consist of the points satisfying (45) together with the points satisfying

$$\alpha < f \leqq b. \tag{47}$$

Thus to put the points satisfying (39) into correspondence with the points satisfying (40) the points satisfying (45) are first made to correspond to themselves. The points satisfying (46) are then put into correspondence with those satisfying (47) by requiring that any point satisfying (46) that arises on the trajectory through it from a value of τ that divides the interval from α to a in a certain ratio, shall correspond to that point satisfying (47) that lies on the same trajectory and arises from a value of τ that divides the interval from α to b in the same ratio.

That the latter correspondence is one-to-one follows from the fact that the trajectories concerned pass through each of the points concerned with one and only one trajectory through each point. That this correspon-

dence is continuous follows from the nature of the dependence of the points of the trajectories on τ and on their initial points on (41). Further, the correspondence of the points satisfying (45) with themselves obviously joins up in a continuous manner with the correspondence of the points satisfying (46) with those satisfying (47). Thus the lemma is proved.

Critical points of the type zero

8. **The connectivity numbers of the complex $f \leqq m + e^2$ where m is the absolute minimum of f.** Let e be a positive constant so small that between m and $m + e^2$ there is no constant equal to a critical value of f. Under the boundary conditions α, § 4, f will take on its absolute minimum at one or more points all interior to S. Let P be one such point. It follows from Lemma 4 of § 3 that the points $(x_1, x_2, \cdots, x_n) = (X)$ neighboring P can be transformed in a one-to-one continuous manner into the points $(y_1, y_2, \cdots, y_n) = (Y)$ neighboring the origin in such a manner that

$$f(x_1, x_2, \cdots, x_n) - m = y_1^2 + y_2^2 + \cdots + y_n^2.$$

Thus the points in the neighborhood of P satisfying

$$f(x_1, x_2, \cdots, x_n) \leqq m + e^2,$$

for a sufficiently small positive constant e, will correspond in a one-to-one continuous manner to the points satisfying

$$y_1^2 + y_2^2 + \cdots + y_n^2 \leqq e^2$$

or what might be described as the points interior to and on an $(n-1)$-dimensional hypersphere in an n-dimensional space. Such a set of points would constitute, in the terms of analysis situs, an n-cell and its boundary. With this understood we can state the following lemma.

Lemma 7. *If f takes on its absolute minimum m in S at s distinct points of S, then for a sufficiently small positive constant e the set of points in S satisfying*

$$f(x_1, x_2, \cdots, x_n) \leqq m + e^2$$

constitutes s distinct n-cells and their boundaries.

From this lemma and the definitions of the connectivity numbers* we have the following theorem.

* V, loc. cit.

THEOREM 1. *If f takes on its absolute minimum m in S at s distinct points of S then for a sufficiently small positive constant e the set of points in S satisfying*

$$f(x_1, x_2, \cdots, x_n) \leqq m + e^2$$

makes up an n-dimensional complex with the connectivity numbers $R_0 = s$, $R_1 = R_2 = \cdots = R_n = 1$.

9. **A critical point of type zero at which f is not equal to its absolute minimum.**

THEOREM 2. *If a and b $(a < b)$ are any two ordinary values of f in S between which there is just one critical value of f, say c, taken on at p_0 critical points all of the zeroth type, then the connectivity numbers R_i, of the set of points in S satisfying*

$$f(x_1, x_2, \cdots, x_n) \leqq b, \tag{48}$$

differ from the connectivity numbers R_i' of the set of points in S satisfying

$$f(x_1, x_2, \cdots, x_n) \leqq a \tag{49}$$

only in that

$$R_0 = R_0' + p_0.$$

Let P be any one of the p_0 critical points of the zeroth type. If e be a sufficiently small positive constant it follows from Lemma 4, § 3, that the points $(x_1, x_2, \cdots, x_n)$ neighboring P and satisfying

$$f \leqq c + e^2 \tag{50}$$

can be made to correspond in a one-to-one continuous manner to a set of points $(y_1, y_2, \cdots, y_n)$ satisfying

$$y_1^2 + y_2^2 + \cdots + y_n^2 \leq e^2. \tag{51}$$

Now the points satisfying (51) make up an n-cell and its boundary. There will thus be associated with each of the p_0 critical points an n-cell and its boundary consisting of points satisfying (50). For e sufficiently small these n-cells will be distinct from each other and from the set of remaining points, say C_n, satisfying (50).

The set C_n is bounded by points at which $f = c + e^2$, and contains only those critical points which are already contained among the points satisfying (49). With the aid of the orthogonal trajectories as used in the proof of Lemma 6, it is easy to show that C_n can be put into one-to-one continuous

correspondence with the set of points satisfying (49). The connectivity numbers of C_n are accordingly equal to the connectivity numbers R_i' of the complex (49).

If now the p_0 n-cells and their boundaries be added to the complex C_n there will result the entire set of points satisfying (50) and the only connectivity number to be thereby changed will be R_0', which will be increased by p_0. If finally we suppose e originally chosen so small that $c+e^2 < b$, it follows from Lemma 6 that the set of points satisfying (50) can be put into one-to-one correspondence with the set of points satisfying (48). Hence (48) defines a complex with the same connectivity numbers as the complex defined by (50). Retracing the steps it is seen that the relation between the connectivity numbers of the complexes (48) and (49) is as stated.

Critical Points of Types $1, 2, \cdots, n-1$

10. **Outline of the method to be followed.** Throughout this chapter we shall assume that we are dealing with a critical value c taken on by f at just one critical point of the kth type, where k may be any one of the integers $1, 2, \cdots, n-1$. We can and will suppose e to be so small a positive constant that there is no critical value of f between $c-e^2$ and $c+e^2$ or equal to $c-e^2$ or $c+e^2$. It is the aim of this chapter to determine the difference between the connectivity numbers of the complex of points (X) satisfying

$$f \leqq c-e^2 \tag{52}$$

and the complex of points (X) satisfying

$$f \leqq c+e^2. \tag{53}$$

To accomplish this we will first show that the complex (52) can be put into one-to-one continuous correspondence with the complex (53) provided we exclude a properly chosen neighborhood of the given critical point. This will be accomplished with the aid of the orthogonal trajectories already used in § 7. The problem then will be resolved into one of determining the difference between the connectivities of the complex (53) and of the complex (53) with this neighborhood excluded.

11. **Simultaneous reduction of $f(x_1, x_2, \cdots, x_n)$ and the differential equations of the orthogonal trajectories to a simple form.** It follows from Lemma 2, § 2, that, in the neighborhood of a critical point $(x_1^0, x_2^0, \cdots, x_n^0)$ at which $f = c$,

$$f - c = \sum_{ij} b_{ij} (x_i - x_i^0)(x_j - x_j^0) \qquad (i, j = 1, 2, \cdots, n), \tag{54}$$

where each b_{ij} is a continuous function of the x's possessing continuous first partial derivatives, and where we can suppose that $b_{ij} = b_{ji}$. In the following the matrix whose element in the ith row and jth column is given by a symbol such as b_{ij} will itself be denoted be the same symbol b without the subscripts. Conjugates of matrices will be indicated by adding a prime.

LEMMA A. *In the neighborhood of a critical point* $(x_1^0, x_2^0, \cdots, x_n^0) = (X_0)$ *there exists a real transformation of the form*

$$x_i - x_i^0 = \sum_j a_{ij} z_j \qquad (i, j = 1, 2, \cdots, n) \tag{55}$$

such that (54) *becomes*

$$\sum_{ij} b_{ij}(x_i - x_i^0)(x_j - x_j^0) = r_1 z_1^2 + \cdots + r_n z_n^2, \tag{56}$$

where $r_1, r_2, \cdots, r_k$ *are negative constants and* $r_{k+1}, r_{k+2}, \cdots, r_n$ *are positive constants, while in* (55) *each* a_{ij} *is a function of the* z*'s possessing continuous first partial derivatives in the neighborhood of the origin of the space of the* z*'s, and at the origin the matrix equation*

$$a(0, 0, \cdots, 0)\ a'(0, 0, \cdots, 0) = \mathrm{I} \tag{57}$$

holds where I *is the unit matrix.*

Proof. From the results of § 3 it follows that in the neighborhood of the given critical point there exists a transformation

$$x_i - x_i^0 = \sum_j e_{ij} y_j \qquad (i, j = 1, 2, \cdots, n) \tag{58}$$

under which

$$\sum_{ij} b_{ij}(x_i - x_i^0)(x_j - x_j^0) = -y_1^2 - y_2^2 - \cdots - y_k^2 + y_{k+1}^2 + y_{k+2}^2 + \cdots + y_n^2 \tag{59}$$

where each e_{ij} is a function of the y's possessing continuous first partial derivatives in the neighborhood of the origin of the space of the y's. Denote by η_{ij} the value of e_{ij} at the origin of the y's. Suppose that under the transformation

$$x_i = \sum_j \eta_{ij} y_j \qquad (i, j = 1, 2, \cdots, n), \tag{60}$$

$$\sum_i x_i^2 = \sum_{ij} \gamma_{ij} y_i y_j. \tag{61}$$

According to the theory of pairs of quadratic forms the quadratic forms in the right hand members of (59) and (61), by a real non-singular transformation with constant coefficients of the form

$$y_i = \sum_j \delta_{ij} z_j \qquad (i, j = 1, 2, \cdots, n) \tag{62}$$

can be reduced to a pair of forms

$$r_1 z_1^2 + r_2 z_2^2 + \cdots + r_n z_n^2,$$
$$z_1^2 + z_2^2 + \cdots + z_n^2.$$

The transformation obtained by expressing the x's in (58) in terms of the z's in (62) is the one whose existence is affirmed in the lemma. For the matrix of this combined transformation, namely $e\delta$, reduces at the origin to $\eta\delta$, and $\eta\delta$ is the matrix of a transformation under which

$$\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} z_i^2,$$

so that $\eta\delta$ is orthogonal as (57) requires. The remaining assertions of Lemma A follow directly.

LEMMA B. *In the neighborhood of the critical point* (X_0) *there exists a real transformation*

$$x_i - x_i^0 = \sum_j y_{ij} y_j \qquad (i, j = 1, 2, \cdots, n) \tag{63}$$

such that

$$\sum_{ij} b_{ij} (x_i - x_i^0)(x_j - x_j^0) = -y_1^2 - y_2^2 - \cdots - y_k^2 + y_{k+1}^2 + \cdots + y_n^2, \tag{64}$$

while the differential equations of the orthogonal trajectories to the manifolds $f = \text{const.}$ *if taken in the form*

$$\frac{dx_h}{dt} = f_h \qquad (h = 1, 2, \cdots, n) \tag{65}$$

are transformed into the equations

$$\frac{dy_h}{dt} = 2r_h y_h + \sum_{ij} A_{hij} y_i y_j \quad (h, i, j = 1, 2, \cdots, n) \tag{66}$$

where $r_1, r_2, \cdots, r_k$ *are positive constants and* $r_{k+1}, r_{k+2}, \cdots, r_n$ *are negative constants, where each* g_{ij} *and* A_{hij} *is a continuous function of the* y*'s*

in the neighborhood of the origin and the g_{ij} there possess continuous first partial derivatives.

Proof. With the aid of the representation of f given in (54), (65) becomes

$$\frac{dx_h}{dt} = 2\sum_i b_{hi}\,(x_i - x_i^0) + \sum_{ij}\frac{\partial b_{ij}}{\partial x_h}(x_i - x_i^0)\,(x_j - x_j^0). \tag{67}$$

It follows from (C), § 2, that each b_{hi} equals its value, say β_{hi} at (X_0), plus a sum of terms each a product of a difference $(x_j - x_j^0)$ multiplied by a function of the x's continuous in the neighborhood of (X_0). Thus (67) becomes

$$\frac{dx_h}{dt} = 2\sum_i \beta_{hi}\,(x_i - x_i^0) + \sum_{ij} B_{hij}\,(x_i - x_i^0)\,(x_j - x_j^0) \tag{68}$$

where each B_{hij} is a continuous function of the x's in the neighborhood of the origin. The transformation (55) of Lemma A will be applied to (68). Remembering that the value of a_{ij} at the origin is written as α_{ij}, it is seen that (55) may be written

$$x_h - x_h^0 = \sum_i \alpha_{hi} z_i + \sum_{ij} C_{hij}\, z_i\, z_j \tag{69}$$

where each C_{hij} is a continuous function of the z's in the neighborhood of their origin.

The transformation (55) applied to (68) gives

$$\frac{dz_h}{dt} = 2\sum_i p_{hi} z_i + \sum_{ij} D_{hij} z_i z_j, \tag{70}$$

where the elements p_{hi} are constants such that, in matrix notation,

$$p = \alpha^{-1}\beta\,\alpha \tag{71}$$

and each D_{hij} is a continuous function of the z's in the neighborhood of their origin. But according to (57) $\alpha\alpha' = 1$ so that $\alpha' = \alpha^{-1}$. Hence (71) may be written as

$$p = \alpha'\beta\,\alpha. \tag{72}$$

But from (55) and (56) it follows that

$$a'\,b\,a = r. \tag{73}$$

Matrix relation (73) becomes at the origin

$$\alpha'\beta\,\alpha = r \tag{74}$$

so that from (74) and (72) we have

$$p = r.$$

Thus (70) takes the form

$$\frac{dz_h}{dt} = 2r_h z_h + \sum_{ij} D_{hij} z_i z_j. \tag{75}$$

Finally the transformation

$$y_h = \sqrt{|r_h|}\, z_h \qquad (h = 1, 2, \cdots, n) \tag{76}$$

changes the right hand member of (56) to the right hand member of (64) and takes (75) into the form (66). Thus the transformation obtained by combining (55) and (76) so as to eliminate the z's is of the nature required of (63) and the lemma is proved.

12. **The correspondence between the complex $f \leqq c - e^2$ and the complex $f \leqq c + e^2$ with the points (79) excluded.** We are concerned with a critical value c taken on by f at just one critical point (X_0) of the kth type $(0 < k < n)$. Let the points (X) in the neighborhood of (X_0) be subjected to the transformation (63) of Lemma B, § 11, so that in terms of the new variables $(y_1, y_2, \cdots, y_n) = (Y)$

$$f - c = q^2 - p^2 \tag{77}$$

where

$$\begin{aligned} p^2 &= y_1^2 + y_2^2 + \cdots + y_k^2, \\ q^2 &= y_{k+1}^2 \quad + \cdots + y_n^2. \end{aligned} \tag{78}$$

It is the purpose of this section to show that for a proper choice of constants a and $e\,(a > e > 0)$, the complex of points satisfying $f \leqq c - e^2$ can be put into one-to-one continuous correspondence with the complex of points satisfying $f \leqq c + e^2$ provided we exclude from the latter complex those points which satisfy

$$c - e^2 < f \leqq c + e^2, \qquad p^2 < a^2. \tag{79}$$

The existence of this correspondence is affirmed in slightly different terms in the following lemma.

LEMMA 8. *If c be a critical value of f taken on by f at just one critical point (X_0), and if (X_0) is of the type for which $0 < k < n$, then for a proper choice of a and e, $a > e > 0$, the points (X) satisfying*

$$f \leqq c - e^2 \tag{80}$$

can be put into one-to-one continuous correspondence with the set of points consisting of the points (80) *together with the points satisfying*

$$c - e^2 < f \leqq c + e^2, \qquad p^2 \geqq a^2.^* \tag{81}$$

The latter inequality $p^2 \geqq a^2$ has a meaning and is intended to have a meaning only where the variables $y_1, y_2, \cdots, y_n$ are defined (Lemma B), that is, in the neighborhood of the given critical point. To prove this lemma, Lemmas C, D, E, F, and G will first be established. For this purpose the boundary points of (81) will be grouped into the following sets of points: the points on $f = c + e^2$ satisfying

$$f = c + e^2, \qquad p^2 > a^2,^* \tag{82}$$

the points on $f = c - e^2$ satisfying

$$f = c - e^2, \qquad p^2 > a^2,^* \tag{83}$$

the points on $p^2 = a^2$ satisfying

$$p^2 = a^2, \qquad c - e^2 < f \leqq c + e^2, \tag{84}$$

and the points

$$p^2 = a^2, \qquad f = c - e^2. \tag{85}$$

LEMMA C. *There exists a positive constant d so small that when a is chosen less than d, e can then be chosen so small that none of the trajectories orthogonal to the manifolds $f =$* const. *passes through more than one point of* (84) *or its boundary.*

Proof of Lemma C. For e sufficiently small each point of (84) lies arbitrarily near some point satisfying

$$p^2 = q^2 = a^2. \tag{86}$$

Let $(B) = (b_1, b_2, \cdots, b_n)$ be a point satisfying (86). Let $(U) = (u_1, u_2, \cdots, u_n)$ be a point in the space of the y's near (B). The solutions of (66) which pass through (U) when $t = 0$ are representable for t near 0 in the form

$$y_h = Y_h(u_1, u_2, \cdots, u_n, t) \qquad (h = 1, 2, \cdots, n), \tag{87}$$

for which

$$u_h \equiv Y_h(u_1, u_2, \cdots, u_n, 0), \tag{88}$$

* Where the latter condition has no meaning through lack of definition of p it will be supposed satisfied vacuously.

where each Y_h possesses continuous partial derivatives with respect to all of its arguments for (U) neighboring (B) and t neighboring 0.

If now any point (U) neighboring (B) be given, the trajectory passing through (U) for $t = 0$ meets the manifold $p^2 = a^2$ in as many points as there are solutions $(y_1, y_2, \cdots, y_n, t)$ of the equations

$$\begin{aligned} y_h &= Y_h(u_1, u_2, \cdots, u_n, t), \\ p^2 &= a^2. \end{aligned} \tag{89}$$

Equations (89) will be written in the form

$$\begin{aligned} y_h - Y_h(u_1, u_2, \cdots, u_n, t) &= 0, \\ y_1^2 + y_2^2 + \cdots + y_k^2 - a^2 &= 0. \end{aligned} \tag{90}$$

The jacobian J of the left hand members of (90) with respect to $y_1, y_2, \cdots, y_n$ and t for $t = 0$ is obtained by bordering an n-square unit matrix I as follows (Lemma B, § 11):

$$J = \begin{vmatrix} & & & & -2r_1 y_1 + \cdots \\ & & & & -2r_2 y_2 + \cdots \\ & & \mathrm{I} & & \cdots\cdots\cdots \\ & & & & \cdots\cdots\cdots \\ & & & & -2r_n y_n + \cdots \\ 2y_1, & 2y_2, \cdots, 2y_k, & 0, \cdots, 0 & & 0 \end{vmatrix}$$

where the terms omitted in the last column are of the form $A_{ij}\, y_i y_j$, where each A_{ij} is a continuous function of the y's in the neighborhood of the origin. The value of this jacobian when $(U) = (Y) = (b_1, b_2, \cdots, b_n)$ is

$$J = 4\sum_h r_h b_h^2 + \sum_{mij} B_{mij}\, b_m\, b_i\, b_j \qquad \begin{pmatrix} m, i, j = 1, 2, \cdots, n; \\ h \quad = 1, 2, \cdots, k \end{pmatrix} \tag{91}$$

where each B_{mij} is a continuous function of the b's in the neighborhood of the origin.

Recalling that $r_1, r_2, \cdots, r_k$ are all negative constants, let m and M be two positive constants such that

$$m < -4r_h < M \qquad (h = 1, 2, \cdots, k).$$

Then

$$m\sum b_h^2 < -4\sum r_h b_h^2 < M\sum b_h^2. \tag{92}$$

Since the point (B) satisfies (86),

$$a^2 = \sum_h b_h^2.$$

Hence from (92)

$$m a^2 < -4 \sum_h r_h b_h^2 < M a^2 \qquad (h = 1, 2, \cdots, k).$$

That is

$$4 \sum_h r_h b_h^2 = R a^2, \tag{93}$$

where R is a function of $(b_1, b_2, \cdots, b_n)$ such that

$$m < |R| < M. \tag{94}$$

Now the point (B) satisfies $p^2 = q^2 = a^2$, from which it follows that

$$|b_i| \leqq a. \tag{95}$$

Moreover since the functions B_{mij} in (91) are continuous functions of the point (B) for (B) in the neighborhood of the origin, there exists a constant H such that for that neighborhood

$$|B_{mij}| < H. \tag{96}$$

For this neighborhood it follows from (95) and (96) that

$$\sum_{mij} B_{mij} b_m b_i b_j < n^3 H a^3,$$

where n^3 appears since there are not more than n^3 terms $b_m b_i b_j$. Thus

$$\sum_{mij} B_{mij} b_m b_i b_j = T a^3,$$

where T is a function of the point (B) such that in the neighborhood of the origin

$$|T| < n^3 H. \tag{97}$$

From (91), (93), and (97) it follows that

$$J = R a^2 + T a^3.$$

We can and do hereby choose a positive constant d so small that for $0 < a < d$

$$J \neq 0.$$

For this choice of a none of the segments of the trajectories with points all lying sufficiently near (B) pass through more than one point of $p^2 = a^2$ neighboring (B). To show that none of these trajectories return to points of (84) it will be useful to take again that representation of the trajectories in which the parameter τ at any point on a trajectory equals the value of f at that point. For e sufficiently small the points on the trajectories passing through some point in a sufficiently small neighborhood of (B) and for which

$$c - e^2 \leqq \tau \leqq c + e^2$$

will all lie arbitrarily near (B) and hence have at most one intersection with $p^2 = a^2$. But these trajectories no matter how extended will not return to points of (84) or its boundary, for points on the extended trajectories will be points for which

$$\tau > c + e^2 \quad \text{or} \quad \tau < c - e^2$$

and hence will be points at which

$$f > c + e^2 \quad \text{or} \quad f < c - e^2,$$

and hence will not be points satisfying (84) or its boundary.

The following lemma follows from the above proof, and in particular from the non-vanishing of the above jacobian.

LEMMA D. *For a choice of a such that the above jacobian is not zero and for any point (U) sufficiently near some point of*

$$p^2 = q^2 = a^2,$$

the trajectory passing through (U) for $t = 0$ meets the manifold

$$p^2 = a^2$$

in a point (Y) whose coördinates and parameter value t are continuous functions of the coördinates of (U) and are provided with continuous first partial derivatives.

The proof of the following lemma could be given along the same lines as the proof of Lemmas C and D.

LEMMA E. *For a sufficiently small positive constant a, and corresponding to this a, a sufficiently small positive constant e, and for any point (U) sufficiently near some point of*

$$p^2 = q^2 = a^2,$$

the trajectory which passes through (U) *when* $t = 0$ *meets the manifold*

$$p^2 - q^2 = e^2$$

in a point (Y) *whose coördinates and parameter value* t *are continuous functions of the coördinates of* (U) *and are provided with continuous first partial derivatives.*

The following lemma will now be proved:

LEMMA F. *For a sufficiently small positive constant* a, *and corresponding to this* a, *a sufficiently small positive constant* e, *all of the trajectories passing through points of* (84) *pass through points of* (83).

The equations (84) may be written in terms of the point (Y) as

$$-e^2 \leqq p^2 - q^2 < e^2, \qquad p^2 = a^2, \tag{98}$$

or again as

$$p^2 = a^2, \qquad a^2 - e^2 < q^2 \leqq a^2 + e^2, \tag{99}$$

while (83) may be written as

$$p^2 - q^2 = e^2, \qquad p^2 > a^2. \tag{100}$$

To distinguish between a point (Y) on (99) and a point (Y) on (100), any point (Y) on (99) will be written as $(Y) = (U) = (u_1, u_2, \cdots, u_n)$. Now (99) and (100) have as common boundary points the points satisfying

$$p^2 = a^2, \qquad q^2 = a^2 - e^2. \tag{101}$$

Let $(B) = (b_1, b_2, \cdots, b_n)$ be any point on (101). The complete set of points on (99) consists of the points on the following family of straight lines:

$$\begin{aligned} u_1 &= b_1, \\ u_2 &= b_2, \\ &\cdots\cdots, \\ u_k &= b_k, \\ u_{k+1} &= m\, b_{k+1}, \\ &\cdots\cdots, \\ u_n &= m\, b_n, \end{aligned} \tag{102}$$

where m is a parameter varying on each straight line on the interval

$$1 < m \leqq \sqrt{\frac{a^2 + e^2}{a^2 - e^2}}. \tag{103}$$

For $m = 1$, $(U) = (B)$. If m, considered as the single independent variable, vary on its interval (103), (U) will then vary on the straight line (102), and the trajectory through (U), according to Lemma E, will meet $p^2 - q^2 = e^2$ in a point (Y) whose coördinates and parameter value t will be continuous functions of m provided with continuous derivatives with respect to m. (U), (Y), and t, thus depending upon m, satisfy (87) making (87) an indentity in m. The lemma will follow readily if it can be shown that as m increases on its interval, and (Y) depends on m as just described, $p^2 = y_1^2 + y_2^2 + \cdots + y_k^2$, considered as a function of m, increases.

To this end let (87) be differentiated with respect to m. At $m = 1$ and hence $t = 0$ (cf. (66)),

$$\frac{dy_h}{dm} = \left[2r_h y_h + \sum_{ij} A_{hij} y_i y_j\right] \frac{dt}{dm} \quad \begin{pmatrix} h = 1, 2, \cdots, k; \\ i, j = 1, 2, \cdots, n \end{pmatrix}, \tag{104}$$

$$\frac{dy_h}{dm} = b_h + \left[2r_h y_h + \sum_{ij} A_{hij} y_i y_j\right] \frac{dt}{dm} \quad (h = k+1, \cdots, n). \tag{105}$$

If the two members of the equation in (104) which gives dy_h/dm be multiplied by $2y_h$ $(h = 1, 2, \cdots, k)$ and the resulting k equations added, the left hand sum will become dp^2/dm, and if in the right hand sum we set $(Y) = (U) = (B)$ we will have

$$\frac{dp^2}{dm} = 4\left[\sum_h r_h b_h^2 + \cdots\right] \frac{dt}{dm} \quad (h = 1, 2, \cdots, k), \tag{106}$$

where the terms omitted here and in the subsequent equations of this proof are sums of terms of the form

$$B_{hij} b_h b_i b_j,$$

where B_{hij} is a continuous function of all of the b's in the neighborhood of the origin.

Similarly, upon multiplying the two members of the equation in (105) that gives dy_h/dm by $2y_h$ $(h = k+1, \cdots, n)$ and adding the resulting $n-k$ equations we will have at $(Y) = (U) = (B)$, using (101),

$$\frac{dq^2}{dm} = 2(a^2 - e^2) + 4\left[\sum_h r_h b_h^2 + \cdots\right] \frac{dt}{dm} \quad (h = k+1, \cdots, n). \tag{107}$$

Now the coördinates of (Y) considered as the above functions of m satisfy $p^2 - q^2 = e^2$ identically. Thus

$$\frac{dq^2}{dm} = \frac{dp^2}{dm}. \tag{108}$$

Upon eliminating $\frac{dq^2}{dm}$ and $\frac{dt}{dm}$ from (106), (107), and (108), we have

$$\frac{dp^2}{dm} = \frac{2\left[\sum_h r_h b_h^2 + \cdots\right](a^2 - e^2)}{\sum_h r_h b_h^2 - \sum_i r_i b_i^2 + \cdots} \quad \begin{pmatrix} h = 1, 2, \cdots, k; \\ i = k+1, \cdots, n \end{pmatrix}$$

(109)

$$= \frac{2\left[\sum_h r_h b_h^2 + \cdots\right](a^2 - e^2)}{-\sum_j |r_j| b_j^2 + \cdots} \qquad (j = 1, 2, \cdots, n).$$

Now if a be a sufficiently small positive constant and e be less, say, than $a/2$, it can be shown by the methods used at the end of the proof of Lemma C that the expression (109) has the sign of

$$\frac{2\sum_h r_h b_h^2}{-\sum_j |r_j| b_j^2},$$

and hence is positive.

For this choice of a and e, dp^2/dm is then positive at (B), that is, when $m = 1$. For a fixed $(a > 0)$ and for e sufficiently small, the interval (103) will be arbitrarily small. Hence for e sufficiently small dp^2/dm will be positive throughout the interval (103). But for $m = 1$, $p^2 = a^2$. Thus on the interval (103) $p^2 > a^2$, and the lemma follows directly.

The preceding proof requires only a formal alteration to serve as the proof of the following lemma.

LEMMA G. *For a sufficiently small positive constant a and, corresponding to this a, a sufficiently small positive constant e, none of the trajectories passing through points of* (84) *or its boundary will pass through a point of* (82).

Proof of Lemma 8. Let us return to that form of representation of the orthogonal trajectories in which the parameter τ at any point on a trajectory equals the value of f at that point. Let (X) be any interior point of (81). On the trajectory through (X) there will be a first point P preceding (X) in the sense of a decreasing τ which is on the boundary of (81), and a first point Q following (X) in the sense of an increasing τ which is on the boundary of (81). It follows from Lemmas C and G that neither P nor Q can lie on (85). P will lie then either on (83) or else on (84), while Q will lie on (82) or else on (84). There are four possibilities:

(a) P on (84), Q on (84),

(b) P on (84), Q on (82),

(c) P on (83), Q^* on (84),

(d) P on (83), Q on (82).

The combination (a) is impossible according to Lemma C, while combination (b) is impossible according to Lemma G.

Thus P always lies on (83). Further, the points P will include each point P_1 on (83). For there are no boundary points of (81) other than points $f = c - e^2$ neighboring P_1. Hence all of the points neighboring P_1 for which $f > c - e^2$ will be interior points of (81). Hence on a trajectory through P_1 the points neighboring P_1 for which $\tau > c - e^2$ will be interior points of (81). Thus P_1 will be one of the points P.

According to (c) and (d) all of the points Q lie either on (82) or else on (84). Further, the points Q include all of the points of (82), a fact which can be proved after the manner of proof in the preceding paragraph. The points Q also include all the points Q_1 of (84). For according to Lemma F a trajectory which passes through a point Q_1 on (84) in the sense of a decreasing τ will meet (83) in a point, say P_1. Now any point of (83) such as P_1 has just been proved to be a point P. Since there can be no point on (82) or (84) between P_1 and Q_1, Q_1 must be a point Q. Thus the points Q include all of the points of (82) and (84). To sum up, the segments PQ include all of the points of (81) and the boundary of (81) save the points of (85). At each of the points of (85) it is convenient to suppose a point P coincident with a point Q.

Let τ_1 and τ_2 be respectively the values of τ at a point P and corresponding Q. The value of τ_1 is $c - e^2$. The position of Q and value of τ_2 are readily seen to be continuous functions of the position of P. To define the correspondence whose existence is affirmed in Lemma 8 let each trajectory PQ be continued from P in the sense of a decreasing τ to a point R at which $\tau = \tau_0$, where

$$\tau_1 - \tau_0 = \tau_2 - \tau_1.$$

The new trajectory segments RP will consist of points satisfying (80). Now let every point of (80) not on a segment RP correspond to itself, and let every point of (80) with parameter value τ on a segment RP correspond to that point on RPQ with parameter value τ', at which

$$(\tau' - \tau_0) = 2(\tau - \tau_0).$$

The correspondence thereby established is readily seen to be of the nature required in the lemma to be proved.

13. **Incidence relations between the boundary of D_n and the remainder of the complex $f \leq c + e^2$.** Following the general outline

of the method given in § 10 this section is devoted first to proving that the set of points D_n satisfying

$$-e^2 \leqq q^2 - p^2 \leqq e^2, \qquad p^2 \leq a^2, \qquad 0 < e < a, \tag{110}$$

make up an n-cell a_n and its boundary, and secondly to investigating the incidence relations between D_n and the remainder of the complex of points satisfying $f \leqq c + e^2$.

For the moment let p and q in (110) be interpreted as the rectangular cartesian coördinates (pq) of a point in a p, q plane. So interpreted the points (pq) which satisfy (110) consist of all those points lying between the two conjugate hyperbolas

$$p^2 - q^2 = e^2, \qquad q^2 - p^2 = e^2$$

which also lie between the two straight lines

$$p = a, \qquad p = -a.$$

Since $a > e$ these two straight lines meet the two conjugate hyperbolas altogether in eight real points. The two-dimensional set of points in the (pq) plane satisfying (110) will be denoted by D_2. D_2 contains the origin and is symmetrical with respect to it.

We will now consider a transformation T of the coördinates (pq) into rectangular coördinates (uv) which carries D_2 into a rectangle E_2 with vertices at the intersections of the straight lines $p = \pm a$ with the hyperbola $q^2 - p^2 = e^2$. By the methods of elementary analytic geometry it is seen that a straight line segment issuing from the origin in the (pq) plane in any direction whatsoever meets the boundary of D_2 in just one point whose distance from the origin varies continuously with the angle the given line segment makes with the p axis. The same is true of the boundary of the rectangle E_2. D_2 and E_2 can be put into a one-to-one continuous correspondence T by requiring that a point (pq) on D_2 that divides the straight line segment from the origin to the boundary of D_2 in a particular ratio shall correspond on E_2 to a point on the same straight line and so placed as to divide the line segment from the origin to the boundary of E_2 in the same ratio.

The following properties of the transformation T will be used.

(A) Under the transformation T the point (pq) is related to its correspondent (uv) as follows:

$$\begin{aligned} u &= r(p^2 q^2)\, p, \\ v &= r(p^2 q^2)\, q, \qquad\qquad r(00) = 0, \end{aligned} \tag{111}$$

where $r(p^2 q^2)$ is a positive, single-valued, continuous function of p^2 and q^2 for each pair (pq) in D_2 not equal to (00). The pair (pq) is obviously also related to its correspondent (uv) as follows:

$$\begin{aligned} p &= \sigma(u^2 v^2)\, u, \\ q &= \sigma(u^2 v^2)\, v, \qquad \sigma(00) = 0, \end{aligned} \tag{112}$$

where $\sigma(u^2 v^2)$ is a positive, single-valued, continuous function of u^2 and v^2 for each pair (uv) in E_2 not equal to (00).

(B) Under T the points (pq) on D_2, that is the points satisfying

$$-e^2 \leqq q^2 - p^2 \leqq e^2, \qquad p^2 \leqq a^2, \tag{113}$$

correspond to the points (uv) on E_2, that is the points satisfying

$$u^2 \leqq a^2, \qquad v^2 \leqq b^2, \tag{114}$$

where we have set $a^2 + e^2 = b^2$.

(C) The points on the boundary of D_2 satisfying

$$p^2 - q^2 = e^2, \qquad q^2 \leqq a^2 - e^2, \tag{115}$$

taken with the points on the boundary of D_2 which satisfy

$$p^2 = a^2, \qquad a^2 - e^2 < q^2 \leqq a^2 + e^2, \tag{116}$$

form two continuous curve segments which correspond under T to those two sides of the rectangle E_2 which satisfy

$$u^2 = a^2, \qquad v^2 \leqq b^2. \tag{117}$$

The points (pq) satisfying (115) are the two finite curve segments cut off from the hyperbola $p^2 - q^2 = e^2$ by the straight lines $p = \pm a$, while the points (pq) satisfying (116) make up the four finite segments of the straight lines $p = \pm a$ which lie between the two conjugate hyperbolas.

(D) The points on the bonndary of D_2 which satisfy

$$q^2 - p^2 = e^2, \qquad p^2 < a^2, \tag{118}$$

correspond under T to those two sides of the rectangle E_2 which satisfy

$$u^2 = b^2, \qquad v^2 < a^2. \tag{119}$$

The points (pq) satisfying (118) obviously make up those two finite segments of the hyperbola $q^2 - p^2 = e^2$ which lie between the two straight lines $p = \pm a$.

The results just obtained properly interpreted will give us the following lemma. In this lemma we will again understand by p^2 and q^2 the expressions

$$\begin{aligned} p^2 &= y_1^2 + y_2^2 + \cdots + y_k^2, \\ q^2 &= y_{k+1}^2 + y_{k+2}^2 + \cdots + y_n^2, \end{aligned} \tag{120}$$

and in addition, in terms of the variables $(z_1, z_2, \cdots, z_n) = (Z)$ we set

$$\begin{aligned} u^2 &= z_1^2 + z_2^2 + \cdots + z_k^2, \\ v^2 &= z_{k+1}^2 + z_{k+2}^2 + \cdots + z_n^2. \end{aligned} \tag{121}$$

LEMMA 9. *There exists a one-to-one continuous correspondence between a set of points* $(y_1, y_2, \cdots, y_n) = (Y)$ *and a set of points* $(z_1, z_2, \cdots, z_n) = (Z)$ (a) *under which the set of points* (Y) *satisfying*

$$-e^2 \leqq q^2 - p^2 \leq e^2, \qquad p^2 \leqq a^2 \tag{122}$$

corresponds to the points (Z) *satisfying*

$$u^2 \leqq a^2, \qquad v^2 \leqq b^2, \tag{123}$$

(b) *while that part of the boundary of* (122) *that consists of points* (Y) *satisfying*

$$p^2 - q^2 = e^2, \qquad q^2 \leqq a^2 - e^2 \tag{124}$$

and of points (Y) *satisfying*

$$p^2 = a^2, \qquad a^2 - e^2 < q^2 \leqq a^2 + e^2 \tag{125}$$

corresponds to those boundary points (Z) *of* (123) *which satisfy*

$$u^2 = a^2, \qquad v^2 \leqq b^2, \tag{126}$$

(c) *while the boundary points* (Y) *of* (122) *satisfying*

$$q^2 - p^2 = e^2, \qquad p^2 < a^2 \tag{127}$$

correspond to those boundary points (Z) *of* (123) *which satisfy*

128) $$v^2 = b^2, \qquad u^2 < a^2.$$

The correspondence whose existence is affirmed in this lemma may be given in terms of the function $r(p^2q^2)$ of (111) as follows:

(129) $$z_i = r(p^2q^2)y_i \qquad (i = 1, 2, \cdots, n),$$

where p^2 and q^2 here as in the future stand for the functions of the y's to which they are set equal in (120).

From (129) we have at once that

(130) $$\begin{aligned} [z_1^2 + z_2^2 + \cdots + z_k^2] &= r^2(p^2q^2)[y_1^2 + y_2^2 + \cdots + y_k^2], \\ [z_{k+1}^2 \quad + \cdots + z_n^2] &= r^2(p^2q^2)[y_{k+1}^2 \quad + \cdots + y_n^2]. \end{aligned}$$

Equations (130) show that the functions p^2 and q^2 are related under the transformation (129) to the functions u^2 and v^2 by equations obtained by squaring the different members of (111). But from equations (111) so squared follow those parts of the results of (B), (C) and (D) which describe in detail how p^2 and q^2 are transformed into u^2 and v^2, and the analytical results there contained abstracted from their geometric setting lead directly to the results (a), (b) and (c) of this lemma.

The correspondence is obviously continuous. That it is one-to-one will now be proved by giving its inverse. From (111) and (112) it follows that if neither (pq) nor $(uv) = (00)$

$$\frac{1}{r(p^2q^2)} = \sigma(u^2v^2).$$

Thus (129) may be written

(131) $$y_i = \sigma(u^2v^2)z_i \qquad (i = 1, 2, \cdots, n),$$

where here u^2 and v^2 stand for those functions of the z's to which they are set equal in (121). Further, (131) obviously also holds for the points just excepted, namely $(Y) = (Z) = (0, 0, \cdots, 0)$. Thus the correspondence is one-to-one, and the lemma is completely proved.

The set of points (Z) satisfying (123) can readily be shown to make up an n-cell and its boundary. Hence the corresponding set of points (Y) satisfying (122) make up an n-cell and its boundary. *The set of points* (Y) *satisfying* (122) *has been termed the set* D_n. D_n will also be used to

designate the set of points $(X) = (x_1, x_2, \cdots, x_n)$ corresponding to the points (Y) of D_n.

Let c be the value of f at the given critical point. All of the points (X) of D_n satisfy

$$f \leqq c + \epsilon^2, \tag{132}$$

and are separated from the remaining points (X) satisfying (132) by those boundary points of D_n satisfying (124) and (125). These boundary points of D_n satisfying (124) and (125) will be termed *the old boundary points of* D_n while the remaining boundary points of D_n, namely those satisfying (127), will be termed *the new boundary points of* D_n. The set of all points of D_n other than the old boundary points of D_n will be called *the new points of* D_n.

The points (Z) satisfying (123) may be represented by a pair of points (PQ) of which P is a point in a k-dimensional space with coördinates $(z_1, z_2, \cdots, z_k)$ satisfying

$$u^2 \leqq a^2,$$

and Q is a point in an $(n-k)$-dimensional space with coördinates $(z_{k+1}, z_{k+2}, \cdots, z_n)$ satisfying

$$v^2 \leqq b^2.$$

Let the $(k-1)$-dimensional hypersphere $u^2 = a^2$ in the space of the points P be denoted by S_{k-1} and the $(n-k-1)$-dimensional hypersphere $v^2 = b^2$ in the space of the points Q be denoted by S_{n-k-1}. With the above convention the results of Lemma 9 may be translated into the following lemma.

LEMMA 10. *The set D_n of points (Y) satisfying* (122) *make up an n-cell and its boundary. It can be put into one-to-one continuous correspondence with the set of pairs of points (PQ) obtained by combining an arbitrary point P interior to or on a $(k-1)$-dimensional hypersphere S_{k-1} with an arbitrary point Q interior to or on an $(n-k-1)$-dimensional hypersphere S_{n-k-1}. In this correspondence, points on the old boundary of D_n correspond to those pairs (PQ) that are obtained by combining an arbitrary point P on S_{k-1} with an arbitrary point Q interior to or on S_{n-k-1}, while points on the new boundary of D_n correspond to those pairs of (PQ) that are obtained by combining an arbitrary point P interior to S_{k-1} with an arbitrary point Q on S_{n-k-1}.*

If we make use of the terms of the preceding lemma, Lemma 8 may be restated and completed as follows.

LEMMA 11. *The complex of points satisfying* $f \leqq c - \epsilon^2$ *can be put into one-to-one continuous correspondence with the complex of points, say* C_n', *consisting of the points satisfying* $f \leqq c + \epsilon^2$ *with the new points of* D_n *excluded. The remaining points of* D_n, *namely the points on the old boundary of* D_n, *are on the boundary of* C_n'.

14. **A representation of D_n in terms of the cells of S_{k-1} and S_{n-k-1}.** Suppose the hypersphere S_{k-1} "covered just once" by a non-singular complex whose matrices of incidence $\mathbf{H}_i$ are given by the equations

$$\mathbf{H}_i = \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} \qquad (i = 1, 2, \cdots, k-1).$$

Denote the two i-cells of S_{k-1} by a_i' and a_i'' $(i = 0, 1, \cdots, k-1)$. Denote the k-cell consisting of the points interior to S_{k-1} by a_k. Similarly suppose S_{n-k-1} covered just once by a non-singular complex whose matrices of incidence $\mathbf{H}_j$ $(j = 1, 2, \cdots, n-k-1)$ are each identical with the preceding matrices $\mathbf{H}_i$. Denote the two j-cells of S_{n-k-1} by b_j' and b_j'' $(j = 0, 1, \cdots, n-k-1)$. Denote the $(n-k)$-cell consisting of the points interior to S_{n-k-1} by b_{n-k}.

Let α be any one of the cells a_i', a_i'' or a_k, and β be any one of the cells b_j', b_j'', or b_{n-k}. If an arbitrary point P of a cell α be combined with an arbitrary point Q of a cell β there will result a set of pairs of points (PQ) which may be considered as the elements of an $(i+j)$-cell. Denote such an $(i+j)$-cell by $(\alpha\beta)$. According to Lemma 10 the complex consisting of all such $(i+j)$-cells will be a non-singular complex covering D_n just once. This complex of $(i+j)$-cells will be adopted as the future representation of D_n.

It follows from Lemma 10 that in the representation of D_n just adopted the interior points of D_n will be represented by the n-cell $(a_k b_{n-k})$ while the new boundary of D_n will be represented by the $(k+j)$-cells $(a_k b_j')$ and $(a_k b_j'')$. Here k is the number specifying the type of critical point being considered, and j takes on the values $0, 1, \cdots, n-k-1$. It should be noted that for each value of j there are just two of these $(k+j)$-cells. The cells $(a_k b_j')$, $(a_k b_j'')$, and $(a_k b_{n-k})$ will be termed *the new cells* of D_n and the remaining cells of D_n *the old cells* of D_n.

The above representation of D_n shows that *there are no new cells of dimensionality less than* k. The points on the boundary of a cell $(\alpha\beta)$ will be represented by the pairs (PQ) obtained by combining a point P on the boundary of α with a point Q on β or its boundary, or by combining a point Q on the boundary of β with a point P on α or its boundary. For the work to follow it is important to determine the $(m-1)$-cells on

the boundary of each new m-cell. We begin with the new cells of lowest dimensionality.

The only $(k-1)$-cells on the boundary of the new k-cell $(a_k b_0')$ will be obtained by combining an arbitrary point of a boundary $(k-1)$-cell of a_k, namely a_{k-1}' or a_{k-1}'', with b_0' giving thus the two $(k-1)$-cells

$$(a_{k-1}' b_0'),\ (a_{k-1}'' b_0') \quad \text{on} \quad (a_k b_0'),$$

both of which are old cells of D_n. Similarly the only $(k-1)$-cells on the boundary of $(a_k b_0'')$ are the old cells

$$(a_{k-1}' b_0''),\ (a_{k-1}'' b_0'') \quad \text{on} \quad (a_k b_0'').$$

For $j = 1, 2, \cdots, n-k-1$ the $(k+j-1)$-cells on the boundary of the new $(k+j)$-cell $(a_k b_j')$ are obtained by combining a boundary $(k-1)$-cell of a_k, namely a_{k-1}' or a_{k-1}'', with b_j', or by combining a boundary $(j-1)$-cell of b_j', namely b_{j-1}' or b_{j-1}'', with a_k, giving thus the $(k+j-1)$-cells

$$(a_{k-1}' b_j'),\ (a_{k-1}'' b_j'),\ (a_k b_{j-1}'),\ (a_k b_{j-1}'') \text{ on } (a_k b_j') \quad (j = 1, 2, \cdots, n-k-1),$$

of which the first two cells are old cells and the latter two cells are new cells. Similarly the $(k+j-1)$-cells on the boundary of the new $(k+j)$-cell $(a_k b_j'')$ are the cells

$$(a_{k-1}' b_j''),\ (a_{k-1}'' b_j''),\ (a_k b_{j-1}'),\ (a_k b_{j-1}'') \text{ on } (a_k b_j'') \quad (j = 1, 2, \cdots, n-k-1),$$

of which the first two are old cells and the latter two are new cells. Finally the $(n-1)$-cells on the boundary of the new n-cell $(a_k b_{n-k})$ are

$$(a_{k-1}' b_{n-k}),\ (a_{k-1}'' b_{n-k}),\ (a_k b_{n-k-1}'),\ (a_k b_{n-k-1}'') \quad \text{on} \quad (a_k b_{n-k}),$$

of which the first two are old cells and the latter two are new cells.

15. **Differences between the connectivity numbers of the complex $f \leqq c-e^2$ and the complex $f \leqq c+e^2$.** It follows from Lemma 11 that the complex $f \leqq c-e^2$ has the same connectivity numbers as the complex C_n' consisting of the points satisfying $f \leqq c+e^2$ with the new points of D_n excluded. The problem of this section then resolves itself into one of determining the relations between the connectivity numbers of the complex C_n' and those of the complex $f \leqq c+e^2$. Denote the latter complex by C_n.

If the new points of D_n be added to C_n' there results C_n. The question then is what is the effect upon the connectivity numbers of adding the

new points of D_n to C_n'? According to Lemma 11 the only points of D_n coinciding with points of C_n' are the old boundary points of D_n. Because of the simple nature of the functions defining the boundaries of these complexes we can and will suppose C_n' broken up into its component cells in such a manner that points of C_n' that coincide with points of D_n all lie on cells of C_n' that coincide with cells of D_n.

Let $\mathbf{H}_i$ $(i = 1, 2, \cdots, n)$ be the ith matrix of incidence of C_n. Let ϱ_i be the rank (mod 2) of $\mathbf{H}_i$. For the same values of i let $\mathbf{H}_i'$ be the ith matrix of incidence of C_n'. Let ϱ_i' be the rank (mod 2) of $\mathbf{H}_i'$. Let $R_j (j = 0, 1, \cdots, n)$ be the jth connectivity number of C_n and R_j' that of C_n'. For the same values of j let α_j be the number of j-cells on C_n and α_j' the corresponding number for C_n'. The following well known equations will be used:

$$
\begin{aligned}
R_0 \quad -1 &= \alpha_0 \quad -\varrho_1 \quad -1, \\
R_1 \quad -1 &= \alpha_1 \quad -\varrho_1 \quad -\varrho_2, \\
R_2 \quad -1 &= \alpha_2 \quad -\varrho_2 \quad -\varrho_3, \\
&\cdots\cdots\cdots\cdots, \\
R_{n-1} -1 &= \alpha_{n-1} -\varrho_{n-1} -\varrho_n, \\
R_n \quad -1 &= \alpha_n \quad -\varrho_n.
\end{aligned} \tag{133}
$$

Equations (133) refer to C_n. Similar equations hold for C_n' and are obtained by priming each letter in (133). From equations (133) and equations (133) with the primes added there can be obtained by a subtraction of the corresponding members of the respective equalities the equations

$$
\begin{aligned}
R_0 \quad -R_0' \quad &= \alpha_0 \quad -\alpha_0' \quad -(\varrho_1 \quad -\varrho_1'), \\
R_1 \quad -R_1' \quad &= \alpha_1 \quad -\alpha_1' \quad -(\varrho_1 \quad -\varrho_1') \quad -(\varrho_2 - \varrho_2'), \\
R_2 \quad -R_2' \quad &= \alpha_2 \quad -\alpha_2' \quad -(\varrho_2 \quad -\varrho_2') \quad -(\varrho_3 - \varrho_3'), \\
&\cdots\cdots\cdots\cdots, \\
R_{n-1} - R_{n-1}' &= \alpha_{n-1} - \alpha_{n-1}' - (\varrho_{n-1} - \varrho_{n-1}') - (\varrho_n - \varrho_n'), \\
R_n \quad -R_n' \quad &= \alpha_n \quad -\alpha_n' \quad -(\varrho_n \quad -\varrho_n').
\end{aligned} \tag{134}
$$

To apply these equations the relations between the matrices $\mathbf{H}_i$ and $\mathbf{H}_i'$ will now be determined. In the first place we have

$$\mathbf{H}_i = \mathbf{H}_i' \qquad (i = 1, 2, \cdots, k-1),$$

since for these values of i there are no new i-cells (§ 14) so that for these values of i, C_n and C_n' contain the same i-cells. According to the results of § 14 the table of incidence of the $(k-1)$-cells with the k-cells of C_n will take the following form:

	x x $\cdots$ x	$(a_k\, b_0')$	$(a_k\, b_0'')$
y		0	0
y		0	0
.		.	.
.		.	.
.		.	.
y		0	0
$(a_{k-1}'\, b_0')$	$\mathbf{H}_k'$	1	0
$(a_{k-1}''\, b_0')$		1	0
$(a_{k-1}'\, b_0'')$		0	1
$(a_{k-1}''\, b_0'')$		0	1

where the x's represent k-cells which it is not desired to give more explicitly and the y's represent $(k-1)$-cells. The rectangle containing $\mathbf{H}_k'$ is supposed filled out with the elements of the matrix $\mathbf{H}_k'$. The matrix $\mathbf{H}_k$ of course consists of the elements of the above two rectangles combined. It can be simplified as follows. The two new k-cells written above the last two columns are on at least one new $(k+1)$-cell (§ 14). The boundary of this $(k+1)$-cell is a k-circuit. Let the columns, other than the last column, which in the table are written under k-cells that are on this k-circuit, be added to the last column. The last column will thereby be reduced to zero (mod 2) without altering the rank of $\mathbf{H}_k$. Let the third from the last row be added to fourth from the last row. As a result there will be but one element not zero in the last two columns. The rank ϱ_k' of the matrix of elements of $\mathbf{H}_k'$ has by these processes not been altered (mod 2). It is seen that for ϱ_k there are two possibilities depending upon the nature of the resulting matrix in the left hand rectangle, namely either

Case 1, $\varrho_k = \varrho_k'$

or $(0 < k < n)$.

Case 2, $\varrho_k = \varrho_k' + 1$

It may be shown by examples that both of these cases actually occur.

The tables of incidence corresponding to the matrices $\mathbf{H}_{k+j}$ for $j = 1, 2, \cdots, n-k-1$ (§ 14) are given by the first table on page 384. In the special case where $k = n-1$ there is no table of this sort.

In this table conventions similar to those of the previous table are made. The matrix $\mathbf{H}_{k+j}$ is of course the matrix of all the elements included in the four rectangles. It can be simplified as follows. First, exactly as in the case of the preceding table there is a set of columns which added to the last column (mod 2) reduce that column to zeros. If now the next to the last row be added to the three other rows that have

	x	x	$\cdots$	x	$(a_k b'_j)$	$(a_k b''_j)$
y					0	0
y					0	0
.					.	.
.					.	.
.		$\mathbf{H}'_{k+j}$			.	.
y					0	0
$(a'_{k-1} b'_j)$					1	0
$(a''_{k-1} b'_j)$					1	0
$(a'_{k-1} b''_j)$					0	1
$(a''_{k-1} b''_j)$					0	1
$(a_k b'_{j-1})$	0	0	$\cdots$	0	1	1
$(a_k b''_{j-1})$	0	0	$\cdots$	0	1	1

unity in the next to the last column, there will be left (mod 2) in the last two rows and columns only one element not zero and that a 1 in the next to the last row and column. It follows that

$$\varrho_{k+j} = \varrho'_{k+j} + 1 \quad (j = 1, 2, \cdots, n-k-1).$$

Finally, corresponding to the matrix $\mathbf{H}_n$ there is the table

	x	x	$\cdots$	x	$(a_k b_{n-k})$
y					0
y					0
.					.
.					.
.		$\mathbf{H}'_n$			.
y					0
$(a'_{k-1} b_{n-k})$					1
$(a''_{k-1} b_{n-k})$					1
$(a_k b'_{n-k-1})$	0	0	$\cdots$	0	1
$(a_k b''_{n-k-1})$	0	0	$\cdots$	0	1

from which it follows that

$$\varrho_n = \varrho'_n + 1.$$

The number of cells of any particular rank which are contained in C_n but not in C'_n can be read off from the preceding tables. We tabulate the following results:

$$\begin{aligned}
&\alpha_i - \alpha'_i = 0, \quad \varrho_i - \varrho'_i = 0 && (i = 1, 2, \cdots, k-1);\\
&\alpha_i - \alpha'_i = 2, \quad \varrho_i - \varrho'_i = 0 && (i = k \quad \text{Case 1});\\
&\alpha_i - \alpha'_i = 2, \quad \varrho_i - \varrho'_i = 1 && (i = k \quad \text{Case 2});\\
&\alpha_i - \alpha'_i = 2, \quad \varrho_i - \varrho'_i = 1 && (i = k+1, k+2, \cdots, n-1);\\
&\alpha_i - \alpha'_i = 1, \quad \varrho_i - \varrho'_i = 1 && (i = n).
\end{aligned}$$

From these equations and equations (134) it follows for a critical point of the kth type $(0<k<n)$ that

$$R_i = R_i' \qquad (i = 0,\ 1,\ \cdots,\ k-2,\ k+1,\ k+2,\ \cdots,\ n)$$

while either

$$R_{k-1} = R_{k-1}' \quad \text{and} \quad R_k = R_k' + 1 \qquad \text{(Case 1)}$$

or

$$R_{k-1} = R_{k-1}' - 1 \quad \text{and} \quad R_k = R_k' \qquad \text{(Case 2)}.$$

That both Case 1 and Case 2 are possible may be shown by simple examples. If we recall that the complex C_n' has the same connectivity numbers as the complex $f \leqq c - e^2$ (Lemma 11, § 13) we have the following fundamental theorem.

THEOREM 3. *Let c be a critical value assumed by f at just one critical point and that a critical point of a type for which $0<k<n$; then for a sufficiently small positive constant e the connectivity numbers R_i of the complex of points satisfying $f \leqq c + e^2$ differ from the connectivity numbers R_i' of the complex of points satisfying $f \leqq c - e^2$ only in that either*

$$R_k = R_k' + 1$$

or else

$$R_{k-1} = R_{k-1}' - 1.$$

CRITICAL POINTS OF THE nTH TYPE

16. Let there be given a critical value of the given function, say c, taken on by f at a critical point of the nth type, and at no other critical point. Let e be a positive constant so small that there is no critical value, other than c, between $c-e^2$ and $c+e^2$, or equal to $c-e^2$ or $c+e^2$. This section is devoted to proving the following theorem.

THEOREM 4. *If c be a critical value of f taken on at just one critical point, and that of the nth type, then for a sufficiently small positive constant e, the connectivity numbers R_i of the complex of points in S satisfying*

$$f \leqq c + e^2 \tag{135}$$

differ from the connectivity numbers R_i' of the complex of points in S satisfying

$$f \leq c - e^2 \tag{136}$$

only in that

$$R_{n-1} = R_{n-1}' - 1.$$

Before proving this theorem the following lemma will be established.

LEMMA. *For any complex C_n lying in a finite part of euclidean n-space, and bounded by a finite number of distinct closed manifolds on each of*

which f takes on a non-critical constant value the connectivity number R_n equals unity.

While a theorem similar to this could be proved under much more general hypotheses this particular lemma is all that we need here. To prove the lemma observe that in the given euclidean n-space we can choose an $(n-1)$-dimensional hypersphere S_{n-1} so large that it contains C_n entirely in its interior. Let A_n be the complex consisting of the points interior to and on S_{n-1}. A_n can be broken up into component cells in such a manner as to contain C_n as a sub-complex of its cells. Now as is well known, for A_n, $R_n = 1$, or in equivalent terms A_n contains no n-circuits. Hence C_n contains no n-circuits, and its connectivity number R_n must also be unity. Thus the lemma is proved.

In terms of the y's of Lemma 4, § 3, f can be represented in the neighborhood of the given critical point in the form

$$f - c = -y_1^2 - y_2^2 - \cdots - y_n^2.$$

If the points (Y) satisfy

$$0 \leqq y_1^2 + y_2^2 + \cdots + y_n^2 < e^2, \tag{137}$$

the corresponding points (X) satisfy the relations

$$c - e^2 < f \leqq c$$

and thus satisfy (135) but not (136).

Denote the complex (135) by C_n and the complex (136) by C_n'. The points (X) corresponding to the points (Y) satisfying (137) obviously make up an n-cell, say a_n, which belongs to C_n. The boundary of a_n consists of points at which the sum of the squares of the y's equals e^2 and so consists of points at which $f = c - e^2$. The boundary of a_n thus lies on the boundary of C_n'. Let a_n be added to C_n', thereby forming a complex denoted by C_n''. If use be made of the orthogonal trajectories employed for a similar purpose in § 7, it is easy to prove that C_n and C_n'' can be put into one-to-one continuous correspondence and hence have the same connectivity numbers R^i.

The question remains as to what is the relation between the connectivity numbers R_i of C_n'' and those R_i' of C_n'. C_n' and C_n'' differ only in that a_n belongs to C_n'' and not to C_n'. If then C_n'' be supposed represented by the cells of C_n' together with the n-cell a_n, it appears that all the matrices of incidence of C_n' and of C_n'' will be identical except the nth ones. From formulas (134) it follows that the connectivity numbers of C_n' and C_n'' are

the same with the possible exception of the nth and $(n-1)$th connectivity numbers.

Now it follows from the preceding lemma that both R_n and R'_n equal unity so that

$$R_n - R'_n = 0. \tag{138}$$

Let α'_i and α''_i be respectively the numbers of i-cells in C'_n and C''_n, and ϱ'_j and ϱ''_j be respectively the ranks of the jth matrix of incidence of C'_n and C''_n. Since a_n is the only n-cell in C''_n that is not in C'_n we have

$$\alpha''_n - \alpha'_n = 1. \tag{139}$$

The last two equations of (134) become under the present notation

$$R_{n-1} - R'_{n-1} = \alpha''_{n-1} - \alpha'_{n-1} - (\varrho''_{n-1} - \varrho'_{n-1}) - (\varrho''_n - \varrho'_n), \tag{140}$$

$$R_n - R'_n = \alpha''_n - \alpha'_n - (\varrho''_n - \varrho'_n). \tag{141}$$

From (138), (139) and (141) it follows that

$$\varrho''_n - \varrho'_n = 1. \tag{142}$$

Since $\alpha''_{n-1} = \alpha'_{n-1}$ and $\varrho''_{n-1} = \varrho'_{n-1}$ it follows from (140) that

$$R_{n-1} = R'_{n-1} - 1.$$

Thus the theorem is completely proved.

GENERAL THEOREMS CONCERNING THE CRITICAL POINTS

17. The case of a critical value of f taken on at several different critical points. Up to this point only those critical values of f have been considered which are taken on at just one critical point. A first theorem in the general case is the following:

THEOREM 5. *Let c be any critical value of f that is not the absolute minimum of f in S, and m_i be the number of critical points of the ith type at which $f = c$. Let e be a small positive constant. Let R_i and R'_i be respectively the ith connectivity numbers of the complex of points in S satisfying*

$$f \leqq c + e^2 \tag{143}$$

and

$$f \leqq c - e^2. \tag{144}$$

Then for e sufficiently small there exist integers $p_0, p_1, \cdots, p_{n-1}$ and $q_1, q_2, \cdots, q_n$, all positive or zero, such that

$$
\begin{aligned}
m_0 &= p_0, \\
m_1 &= p_1 + q_1, \\
m_2 &= p_2 + q_2, \\
&\cdots\cdots\cdots, \\
m_{n-1} &= p_{n-1} + q_{n-1}, \\
m_n &= q_n,
\end{aligned}
\tag{145}
$$

and such that

$$
\begin{aligned}
R_0 - R_0' &= p_0 - q_1, \\
R_1 - R_1' &= p_1 - q_2, \\
&\cdots\cdots\cdots, \\
R_{n-1} - R_{n-1}' &= p_{n-1} - q_n,
\end{aligned}
\tag{146}
$$

while, as ever,

$$R_n = R_n' = 1. \tag{147}$$

Denote the complex defined by (143) by C_n, and that defined by (144) by C_n'. The critical points at which $f = c$ are divided for convenience into those of type 0, those of type n, and those of the remaining types. Corresponding to this division of the critical points into classes the proof of the above theorem is divided into parts A, B, and C.

A. *Critical points of type* 0 *at which* $f = c$. For e sufficiently small it follows exactly as in the proof of Theorem 2 of § 9 that the points in C_n neighboring the m_0 critical points of type 0 make up m_0 n-cells and their boundaries, distinct from each other and from any other points of C_n, and not included at all among the points of C_n'. If these m_0 n-cells and their boundaries be added to C_n' there will result a complex, say C_n'', whose connectivity numbers R_n'' will differ from those of C_n' only in that

$$R_0'' - R_0' = m_0. \tag{148}$$

B. *Critical points of type* n *at which* $f = c$. For e sufficiently small it follows exactly as in the proof of Theorem 4, § 16, that the points in the neighborhood of the m_n critical points of type n make up m_n distinct n-cells whose boundaries belong to C_n', but which themselves belong only to C_n. If these m_n n-cells be added to C_n'' there will result a complex C_n''' whose connectivity numbers R_i''' will differ from those of C_n'' only in that

$$R_{n-1}''' - R_{n-1}'' = -m_n. \tag{149}$$

C. *Critical points not of type* 0 *or* n *at which* $f = c$. With each critical point of this class there can be associated a complex of points D_n, exactly as is done in Lemmas 10 and 11 for the case where the given

critical point is the only critical point at which $f = c$. For e and a sufficiently small the different complexes D_n will all be distinct from each other. The so called "new points" of each complex D_n will belong to C_n. Suppose that there are just r critical points, at which $f = c$ and which are of a type neither 0 or n. Let A_j be the set of new points in the complex D_n associated with the jth one of these critical points ($j = 1, 2, \cdots, r$). Let all of the points of C_n not already in C_n''' and not among the points of any set A_j be added to C_n''', thereby forming a complex C_n''''. C_n'''' and C_n''' can be put into one-to-one continuous correspondence by the methods of Lemma 11, § 13, so that the complexes C_n'''' and C_n'''' will have the same connectivity numbers.

Finally to obtain C_n we have only to add to C_n'''' the different sets A_j. We will suppose that these sets A_j are added in the order of their subscripts. Suppose that A_j was associated with a critical point of type k_j. The addition of A_j to what we shall call the *old complex*, consisting of C_n'''' and the points of $A_1, A_2, \cdots, A_{j-1}$, will give a *new complex* whose connectivity numbers will differ from the connectivity numbers of the old complex, either (Case 1) in that the k_jth connectivity number of the new complex will be one greater than that of the old complex, or else (Case 2) the (k_j-1)th connectivity number of the new complex will be one less than that of the old complex. We now define p_i, for $i = 1, 2, \cdots, n-1$, as the number of critical points of type i, at which $f = c$, which come under Case 1, and define q_i for the same values of i as the number of critical points of the ith type at which $f = c$ which come under Case 2. The connectivity numbers R_i'''' of C_n'''' are obviously then related to the connectivity numbers R_i of C_n as follows:

$$
\begin{aligned}
R_0 \quad - R_0'''' &= \quad\quad - q_1, \\
R_1 \quad - R_1'''' &= p_1 \quad - q_2, \\
&\cdots\cdots\cdots\cdots, \\
R_{n-2} - R_{n-2}'''' &= p_{n-2} - q_{n-1}, \\
R_{n-1} - R_{n-1}'''' &= p_{n-1}.
\end{aligned} \tag{150}
$$

We now define p_0 as equal to m_0 and define q_n as equal to m_n. If use be made of the fact that the connectivity numbers of C_n' and C_n'' have no differences other than that given by (148), and that the connectivity numbers of C_n'' and C_n''' have no differences other than that given by (149), while $R_i'''' = R_i'''$ without exception, then from equations (150) and the definitions of the p's and the q's the theorem follows.

18. **Relations between the connectivity numbers of the complex $f \leqq c_1$ and the complex $f \leq c_2$.**

THEOREM 6. *Let c_1 and c_2 $(c_1 < c_2)$ be two non-critical values taken on by f in S. Let m_i be the number of critical points of the ith type at which f equals some constant between c_1 and c_2. Let R_i and R_i' be the ith connectivity numbers of the complexes of points in S respectively satisfying $f \leqq c_2$ and $f \leqq c_1$; then there exist integers, $p_0, p_1, \cdots, p_{n-1}$, and $q_1, q_2, \cdots, q_n$, all positive or zero, such that*

$$m_0 = p_0,$$
$$m_i = p_i + q_i \qquad (i = 1, 2, \cdots, n-1),$$
$$m_n = q_n,$$

and such that

$$R_i - R_i' = p_i - q_{i+1} \qquad (i = 0, 1, \cdots, n-1)$$

while as ever

$$R_n = R_n' = 1.$$

Let $a_1, a_2, \cdots, a_r$ be the critical values of f between c_1 and c_2, taken in the order of increasing magnitude. With each of these critical values of f, say a_j, let there be associated a positive constant e_j taken so small that for this choice of e_j Theorem 5 holds if c and e in Theorem 5 are here taken to be a_j and e_j respectively. Let the constants e_j be also taken so small that the constants

$$(151) \quad c_1,\ a_1 - e_1,\ a_1 + e_1,\ a_2 - e_2,\ a_2 + e_2,\ \cdots,\ a_r - e_r,\ a_r + e_r,\ c_2$$

are none of them equal and appear in (151) in the order of increasing magnitude. Now there are no critical values of f between c_1 and $a_1 - e_1$. It follows from Lemma 6, § 7, that the complex $f \leqq c_1$ has the same connectivity numbers as the complex $f \leqq a_1 - e_1$. For the same reason the complexes $f \leqq a_i + e_i$ and $f \leqq a_{i+1} - e_{i+1}$, for $i = 1, 2, \cdots, r-1$, have the same connectivity numbers, as well as the complexes $f \leqq a_r + e_r$ and $f \leqq c_2$. The relations between the connectivity numbers of the complex $f \leqq a_j + e_j$ and the complex $f \leqq a_j - e_j$ are of the nature of those given by the preceding Theorem 5.

If then there be considered in succession the complexes of points in S satisfying $f \leqq c$, where c takes on, in succession, the constants in (151) in their order from left to right, the relations between these successive complexes will combine into the relations of which this theorem affirms the existence.

19. **Relations between all of the critical points in S and the connectivities of S.** The following theorem is proved under the boundary conditions α of § 4.

THEOREM 7. *Let M_i be the total number of critical points of f of the ith type $(i = 0, 1, \cdots, n)$. Let R_i be the jth connectivity number of S.*

Under the boundary conditions α there exist integers $P_0, P_1, \cdots, P_{n-1}$ and $Q_1, Q_2, \cdots, Q_n$, all positive or zero, such that

$$
\begin{aligned}
M_0 &= P_0 + 1, \\
M_i &= P_i + Q_i \qquad (i = 1, 2, \cdots, n-1), \\
M_n &= \quad Q_n
\end{aligned} \tag{152}
$$

and such that

$$R_i - 1 = P_i - Q_{i+1} \qquad (i = 0, 1, \cdots, n-1) \tag{153}$$

while as ever

$$R_n - 1 = 0. \tag{154}$$

Let s be the number of critical points at which f takes on its absolute minimum m. According to Theorem 2, § 9, for a sufficiently small positive constant e, the connectivity numbers R_i' of the complex of points satisfying $f \leqq m + e^2$ are given by the relations

$$R_0' = s, \qquad R_1' = R_2' = \cdots = R_n' = 1. \tag{155}$$

The proof of these results in Theorem 2 presupposed that e was so small that there were no other critical points than the s absolute minimum points of f among the points in S satisfying $f \leqq m + e^2$. This supposition is again made here. For this choice of e, an application can now be made of Theorem 6. In this application the notation of Theorem 6 will be taken over. The constants c_1 and c_2 of Theorem 6 will here be taken respectively as $m + e^2$ and the value M which f takes on upon the boundary of S. Observe now that M_0, the total number of critical points of type 0 in S, equals $s + m_0$. This we write

$$m_0 = M_0 - s. \tag{156}$$

Otherwise

$$m_i = M_i \qquad (i = 1, 2, \cdots, n). \tag{157}$$

We now give the following definitions of $P_0, P_1, \cdots, P_{n-1}$ and of $Q_1, Q_2, \cdots, Q_n$:

$$
\begin{aligned}
P_0 &= p_0 + s - 1, \\
P_i &= p_i \qquad (i = 1, 2, \cdots, n-1), \\
Q_j &= q_j \qquad (j = 1, 2, \cdots, n),
\end{aligned} \tag{158}
$$

and note that P_0, in particular, as defined, is either positive or zero, since p_0 is at least zero, and s at least one.

If in the relations of Theorem 6, for the constants R_i' there be substituted their values as given by (155), and for m_0 and m_i their values from (156)

and (157) and for p_0 and the p_i and q_j their values as given by (158), there will result the equations which are to be proved.

The following theorem gives the answer to one of the fundamental questions with which this paper is concerned.

THEOREM 8. *Under the boundary conditions α the following relations exist between the set of numbers M_i, where M_i is the number of critical points of the ith type of f in S, and the set of connectivity numbers R_j of S $(i, j = 0, 1, \cdots, n)$:*

$$
\begin{aligned}
1 &\leqq (M_0 - R_0 + 1),\\
1 &\geqq (M_0 - R_0 + 1) - (M_1 - R_1 + 1),\\
1 &\leqq (M_0 - R_0 + 1) - (M_1 - R_1 + 1) + (M_2 - R_2 + 1),\\
1 &\geqq (M_0 - R_0 + 1) - (M_1 - R_1 + 1) + (M_2 - R_2 + 1) - (M_3 - R_3 + 1),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,\\
1 &\leqq [(M_0 - R_0 + 1) - + \cdots + (-1)^{n-1}(M_{n-1} - R_{n-1} + 1)](-1)^{n-1},\\
1 &= (M_0 - R_0 + 1) - + \cdots + (-1)^n \quad (M_n \quad - R_n \quad + 1).
\end{aligned}
$$

To prove the theorem consider equations (152) together with equations (153), without the last equation of (152). These equations together form the system

$$
\begin{aligned}
M_0 &= P_0 + 1, &&\\
M_i &= P_i + Q_i && (i = 1, 2, \cdots, n-1),\\
R_j - 1 &= P_j - Q_{j+1} && (j = 0, 1, \cdots, n-1).
\end{aligned} \tag{159}
$$

Suppose the left hand members of (159) known constants, and the P's and Q's in the right hand members unknown constants or variables. The determinant of the coefficients of these unknown variables is readily seen to be different from zero, so that if the M's and the connectivity numbers are known, the equations (159) uniquely determine the P's and Q's as constants satisfying (159). In particular the equations (159) can be solved successively for the Q's as follows:

$$
\begin{aligned}
1 + Q_1 &= (M_0 - R_0 + 1),\\
1 - Q_2 &= (M_0 - R_0 + 1) - (M_1 - R_1 + 1),\\
1 + Q_3 &= (M_0 - R_0 + 1) - (M_1 - R_1 + 1) + (M_2 - R_2 + 1),\\
1 - Q_4 &= (M_0 - R_0 + 1) - (M_1 - R_1 + 1) + (M_2 - R_2 + 1) - (M_3 - R_3 + 1),\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,\\
1 + (-1)^{n-1} Q_n &= (M_0 - R_0 + 1) - + \cdots + (-1)^{n-1}(M_{n-1} - R_{n-1} + 1).
\end{aligned} \tag{160}
$$

Of the equations of Theorem 7 which are not included in (159) there remain the equations

$$
\begin{aligned}
R_n - 1 &= 0,\\
M_n &= Q_n.
\end{aligned} \tag{161}
$$

With the aid of (161) we can write

$$Q_n = M_n - R_n + 1.$$

If this expression for Q_n be substituted for Q_n in the last of equations (160) there will result the equality given by the last of the equations of Theorem 8 The remaining relations of Theorem 8 result immediately from (160) upon recalling that the Q's are all either positive or zero. Thus the theorem is proved.

If the domain S be one which can in particular be put into one-to-one continuous correspondence with the points in n-space interior to and on an $(n-1)$-dimensional hypersphere, then the connectivity numbers R_i of S are all unity. Hence we have the following corollary to Theorem 8.

COROLLARY. *If S be an n-dimensional region homeomorphic with the points in n-space interior to and on an $(n-1)$-dimensional hypersphere, and M_i be the number of critical points of the ith type of f in S $(i = 0, 1, \cdots, n)$ then under the boundary conditions α the following relations hold true:*

$$\begin{aligned}
1 &\leqq M_0, \\
1 &> M_0 - M_1, \\
1 &< M_0 - M_1 + M_2, \\
1 &\geqq M_0 - M_1 + M_2 - M_3, \\
&\cdots\cdots\cdots\cdots \\
1 &\lesseqgtr [M_0 - M_1 + M_2 - M_3 + - \cdots + (-1)^{n-1} M_{n-1}] (-1)^{n-1}, \\
1 &= M_0 - M_1 + M_2 - M_3 + - \cdots + (-1)^n M_n.
\end{aligned}$$

20. **The boundary conditions β defined.** The boundary conditions β are the following.

I. The boundary B of the domain R of § 1 shall consist of a closed set of points lying in a finite part of the space of the variables $(x_1, x_2, \cdots, x_n)$.

II. The points on B in the neighborhood of any particular point (X_0) on B shall satisfy a relation of the sort

$$F(x_1, x_2, \cdots, x_n) = 0,$$

where $F(x_1, x_2, \cdots, x_n)$ is a single-valued continuous function of $(x_1, x_2, \cdots, x_n)$ possessing continuous first, second, and third partial derivatives in the neighborhood of each point of B, of which not all of the first partial derivatives are to be zero at any point of B.

III. The function $f(x_1, x_2, \cdots, x_n)$ in addition to satisfying the conditions of § 1 shall be continuous on B, and its first partial derivatives shall take on continuous boundary values on B.

IV. At each point of B the unilateral directional derivative of $f(x_1, x_2, \cdots, x_n)$ along the normal to B in the sense that leads from points of R to points not in R, and on the side that lies in R, shall be positive.

21. **The boundary conditions β reduced to the boundary conditions α** (§ 4). By a redefinition of $f(x_1, x_2, \cdots, x_n)$ in the neighborhood of B there can be obtained a function $L(x_1, x_2, \cdots, x_n)$ with critical points identical in position and in type with those of $f(x_1, x_2, \cdots, x_n)$, but such that one of the manifolds

$$L(x_1, x_2, \cdots, x_n) = \text{const.}$$

will serve as a boundary A satisfying the earlier boundary conditions.

Under condition II of the boundary conditions β there is a definite normal to B at each point $(x_1^0, x_2^0, \cdots, x_n^0) = (X_0)$ of B whose equations may be given in the form

$$x_i - x_i^0 = A_i(x_1^0, x_2^0, \cdots, x_n^0)s \qquad (i = 1, 2, \cdots, n), \tag{162}$$

where $A_i(x_1^0, x_2^0, \cdots, x_n^0)$ together with its first and second partial derivatives are continuous functions of their arguments in the neighborhood of any given point on B, where s is the distance along the normal, being zero on B, and increasing as the point on the normal crosses B from points of R to points not in R. We can and will suppose s_1 to be a negative constant chosen so small that the points $(x_1, x_2, \cdots, x_n) = (X)$ on the normal (162) at the points at which

$$s_1 \leqq s < 0 \tag{163}$$

include once and only once every point of R, not including B, in the neighborhood of B.

Because of the boundary conditions IV we can and will suppose that s_1 is so small that the points (X) given by (162) for values of s satisfying (163) contain no critical points of f. It follows that there is one and only one of the trajectories orthogonal to the manifolds $f = \text{const.}$, through each point of (162) for which (163) holds. Let s_2, s_3, and s_4 be three constants such that

$$s_1 < s_2 < s_3 < s_4 < 0.$$

Let T_1, T_2, T_3, and T_4 be respectively the $(n-1)$-dimensional manifolds of points (X) obtained by putting $s = s_1$, s_2, s_3, and s_4 in (162) and letting (X_0) vary on B. It is a consequence of boundary conditions IV, as the calculation of the appropriate jacobian will show, that if s_1 be a sufficiently small negative constant, any trajectory orthogonal to the manifolds $f = \text{const.}$ passing through a point P_1 of T_1 also passes through

uniquely determined points P_2, P_3, and P_4 on T_2, T_3, and T_4, respectively, such that the positions of P_2, P_3, and P_4 are continuous functions of the position of P_1 on T_1.

We again make use of that representation of the orthogonal trajectories in which the parameter at any point equals the value of f at that point, and we let τ_1, τ_2, τ_3, and τ_4 respectively be the values of τ at P_1, P_2, P_3 and P_4.

We now are in a position to replace $f(x_1, x_2, \cdots, x_n)$ by a new function $L(x_1, x_2, \cdots, x_n)$ defined as follows. At each point of R not a point given by (162) for a value of $s > s_2$, we define

$$L(x_1, x_2, \cdots, x_n) \equiv f(x_1, x_2, \cdots, x_n). \tag{164}$$

To define $L(x_1, x_2, \cdots, x_n)$ further it will be convenient, for those points (X) given by (162) for which

$$s_1 \leqq s \leqq s_4, \tag{165}$$

to replace $(x_1, x_2, \cdots, x_n)$ by a new set of n independent variables, namely $(n-1)$ independent parameters $(u_1, u_2, \cdots, u_{n-1})$, determining a point P_1 on T_1 and thereby a trajectory through P_1, and an nth independent parameter, namely the above parameter τ determining a point (X) on the trajectory through P_1. To represent a set of points (X) satisfying (165) and neighboring a particular trajectory, the parameters $(u_1, u_2, \cdots, u_{n-1})$ can be so chosen that the coördinates of (X) are single-valued functions of the variables $(u_1, u_2, \cdots, u_{n-1}, \tau)$ provided with continuous second partial derivatives. Now in terms of the new variables the points satisfying (165) are the points at which

$$\tau_1 \leqq \tau \leqq \tau_4. \tag{166}$$

Of these points the points at which

$$\tau_1 \leqq \tau \leqq \tau_2 \tag{167}$$

are points at which $L(x_1, x_2, \cdots, x_n)$ has already been defined. Setting

$$L(x_1, x_2, \cdots, x_n) = \varphi(\tau, u_1, u_2, \cdots, u_{n-1})$$

the previous definition of L, namely (164), reduces for the points of (167) to

$$\varphi(\tau, u_1, u_2, \cdots, u_{n-1}) \equiv \tau. \tag{168}$$

To define $L(x_1, x_2, \cdots, x_n)$ further note that τ_1, τ_2, τ_3, and τ_4 are functions of $(u_1, u_2, \cdots, u_{n-1})$. Let G be a constant greater than the

value of f in R. For those points of (166) for which

$$\tau_2 \leqq \tau \leqq \tau_4 \tag{169}$$

we now define

$$\varphi(\tau, u_1, u_2, \cdots, u_{n-1}) \equiv \tau + H(u_1, \cdots, u_{n-1})[\tau - \tau_2(u_1, \cdots, u_{n-1})]^4 \tag{170}$$

where H is defined by the equation

$$G = \tau_3 + H[\tau_3 - \tau_2]^4 \tag{171}$$

as a positive function of $(u_1, u_2, \cdots, u_{n-1})$. From (170) and (171) it follows that

$$\varphi(\tau_3, u_1, u_2, \cdots, u_{n-1}) = G. \tag{172}$$

For any point corresponding to a value of τ satisfying (169) we now define

$$L(x_1, x_2, \cdots, x_n) \equiv \varphi(\tau, u_1, \cdots, u_{n-1})$$

at that point.

Let R' be the domain in which L has been defined. $L(x_1, x_2, \cdots, x_n)$ and its second partial derivatives are readily seen to be continuous in R'. No critical points are introduced by the definition (170), as follows upon verification of the fact that the partial derivative $\varphi_\tau(\tau, u_1, \cdots, u_{n-1}) > 0$ for points for which (170) and (171) hold. Thus in the neighborhood of *critical points* of f or of L, (164) holds. From (172) it follows that on the manifold T_3

$$L(x_1, x_2, \cdots, x_n) = G.$$

Boundary conditions α of § 4 will obviously be satisfied if we take for the manifold A used in boundary conditions α the manifold T_3. Further, with the aid of the normals (162) it is easy to see that the complex of points satisfying $L(x_1, x_2, \cdots, x_n) \leqq G$ can be put into one-to-one continuous correspondence with the complex consisting of the domain R and the above boundary B. Thus these two complexes have the same connectivity numbers. From Theorem 8 we obtain accordingly the following theorem.

THEOREM 9. *Under the boundary conditions β the relations between the numbers M_i, where M_i $(i = 0, 1, \cdots, n)$ is the number of critical points of f of type i in R, and the numbers R_j, where R_j $(j = 0, 1, \cdots, n)$ is the jth connectivity number of the complex consisting of R and its boundary B, are the same as those given in Theorem* 8.

CORNELL UNIVERSITY,
ITHACA, N. Y.